DOI: 10.21917/ijsc.2020.0319

PREDICTION OF CROP GROWTH USING MACHINE LEARNING BASED ON SEED FEATURES

N. Nandhini and J. Gowri Shankar

Department of Computer Science and Engineering, Knowledge Institute of Technology, India

Abstract

The presence of plant species in the wrong place and time is identified as plant weeds. The loss of yield could result from interference with plant crop weed species. To classify the plant weeds among the seeds, seed classification is carried out in this paper. Here the image of seeds and datasets of sample seed are input. In pre-processing stage the seed image is given as input. An unwanted seeds are removed by comparing the features of seed with sample seeds features by using ID3 algorithm. One of the reasons for failure in crop yield production is selecting suitable soil for crop. As sample dataset contains the detail of growth of crop in soil, it will help for selecting the suitable soil for seeds. The features of sample dataset are compared with the features extracted from the affected crop and predict the disease and prevention measures taken place. In this method prediction is done only after the growth of crop which leads to decrease in quality of crop growth. In order to overcome these issues, diseases can be predicted using seed features by comparing the features with sample dataset. Using Support Vector Machine algorithm the seeds are classified based on the growth and predicting the diseases of crop. This is done by training the dataset by comparing the features extracted from new seeds and features of sample seeds and predicting the crop growth and diseases. Based on prediction of crop growth and crop diseases, preventive measures takes place.

Keywords:

Data Mining, Big Data, Feature Extraction, Crop Growth

1. INTRODUCTION

Machine Learning is a computer science field where new developments have recently been developed, which also helps to automate assessment and processing carried out by mankind so that human manual power is reduced. Machine learning is a type of artificial intelligence (AI) that enables computers to learn without explicitly programming, according to the technological goal.

Machine learning focuses on computer program, which can change if new data are exposed. It is difficult for novice farmers to find the right crops based on the appearance of the soil Agricultural decline must also be prevented. For ensuring a food security, an effective use of agricultural land is crucial. In this document, we propose an ID3 Algorithm and Support Vector Machine (SVM) crop recommendation system.

This paper describes machine learning, tasks involved, models, advantages, and the purpose of the proposed work and the overview of a thesis. This paper describes an ID3 algorithm and a support vector machine crop prediction and recommendation system. In order to achieve the predicted cropping return, the classified crop field images along with the historical weather and yield data are modelled and recommended for a specific field appropriate crops. This in turn involves classification and prediction machine learning and image processing.

2. RELATED WORKS

The plant weeds identified by Kurtulmus et al. [6] as the presence in the wrong place and time of the plant species. Up to 90% yield loss may be caused by interference by weed species with plant crops.

Arya and Semwal [9] describe how the image is denoted and compares the filters. The processing of images can apply to the input data a much broader range of algorithms, and can prevent problems like noise build-up and signal distortion during image processing. Transforming wavelets has become a very potent tool to de-noise an image. Wiener filter is one of the most common methods. He indicated that the median filter, the median filter, and the Wiener filter use four types of noise and de-noise images for different noises. In comparison to average and medium filter, Viennese filters are the best filters for using the removing noise. The Viennese filter is better than other filters in the removal of noise. Median filter also delivers better noise removal results. As comparison parameters, PSNR MSE and RMSE were used.

Tujo et al. [13] tried to develop a predictive model for predicting seed classes using algorithms for machine learning which result in high crop production. This research machine learning algorithm will be used for development to learn from information that can be used to make predictions, to create real-world simulations, to identify patterns and to classify the input data. A neural artificial network is used to model and find patterns in data for complex relationships between inputs and outputs. This thesis aims to understand the machine-learning algorithm through neural networks and build models that predict machine-learning seed classes. The model developed is experimented with a seed dataset and then seed classes are forecast using the model developed. Finally, the factors determining seed classifications are identified and ranked [5] using the model developed.

In [2], the author states that soil samples using different scientific features are classified and graded on an SVM basis. To acquire and process the coloured pictures of the soil samples, different algorithms and filters are developed. These developed algorithms are used for extracting different characteristics such as colour and texture.

Ghosh and Koley [11] describes an analysis of key soil characteristics such as organic matter, essential plant nutrients, micronutrients that affect plant growth, as well as the percentage that is appropriate for such properties using Supervised Learning, Neural Network for Back Propagation. These parameters can be directly measured, but their measurement is difficult and costly. In both cases, BPN will find, and propose the correct correlation percentage between these properties, reference crop-growth properties and the ability of the NPN to supplies nutrients in their own reserves and through external applications for crop production.

Joshi et al. [4] describes the early detection mechanism for agricultural crop deletion, using machine learning and image processing techniques [3], leaves illnesses for quantitative and qualitative product safety. Architecture of the net. Net. The models work better for the coloured version of the dataset [4] in the case of the dataset used (colour, grey and segmented). An image processing technique can detect the diseased parts of the plant leaves. They researched banana, bean, guavan, jackfruit, lemon, mango, potato, sapota and tomato leaf diseases and classified images in any healthy or one class of diseases. They examined them in the following categories: To achieve this, 500 leaves were used with digital camera from 30 different native plant species. The images obtained have been converted to HSI. Segmentation of the infected part of each leaf was used to extract. As image characteristics play an important role in the classification and training, texture characteristics were removed using the colour co-occurrence method. Because early squatter, brown spots, late squash, fungal and bacterial diseases, SVM (94.74 % with accuracy) and minimum distance criterion (86.77 % accuracy) have been detected [9] [10].

Kima et al. [15] describes the trends in crop pest prediction using machine learning. The field of agriculture also concentrates on it with the advent of data mining. Different studies at home and abroad are currently under way using machine learning technology, and their uses are on the rise. The MSM (Support Vector Machine) classification, multiple linear regression, neural network and Bayesian network technology are introduced and certain cases of their use are described. Their paper analysed the meteorological factors to predict the moisture of the leaves and has developed a GNR for moisture prediction.

Automation of crop disease diagnosis, in particular for regions with few experts, is an important task by Owomugisha et al. [7]. Much of the current methods detect diseases through the analysis of leaf images, in particular for aerial diseases. In order to develop a good classifier one requires an enormous image dataset and proper methods to unambiguously extract relevant image functions from the illness. Image data are also likely to occlude effects that make it hard to analyse the data consistently. We will examine the use of spectral data from leaves of a plant in this paper. We analysis spectral information from visibly illnessed parts of a leaf and visibly healthy parts. In a three classification problem configuration, we employ prototype based methods of classification and standard classification models.

Corrales et al. [3] develops a regression and classification model called controlled education. The purpose of such algorithms is to predict or detect new data input through example. Researchers have inspired the idea to use information as examples by applying supervised learning algorithms to predict future diseases and plagues in agricultural crops. Climate change has threatened farming production; extreme temperatures and humidity, as well as other abiotic stresses, contribute to the ethology of disease and pesticide on crops.

Recent research focused on the prediction and use of techniques like supervised learning algorithms for diseases and pesticides This paper provides an overview of the learning monitoring algorithms commonly employed in agriculture to detect pests and diseases in crops like maize, rice, coffee, mango, peanut, or tomatoes, with the objective of selecting the best algorithms for the agricultural sector.

In order to find efficient crop yield estimates, Vats et al. [14] uses descriptive analyse which predicts the future in farm production. In this paper we've got 3 datasets, such as Soil, Rainfall and Yield. And we produce a combined dataset and use several supervised techniques on this combined dataset to find the actual estimated cost of several techniques. The additional functionality in LS-SVM in comparison with Support Vector Maker was explained. This technique is used to solve problems with classification and regression.

When the data set is large and complex, we typically use a linear version of the vector support machine in comparison with the support vector. The reason basically is that if we are solving a classification or regression problem through the use of support of the vector machine and the data set, the implementation logic for SVM will solve this problem or classify the data points in quadratic time. Whereas the vector support machine solves the problem in linear time in Linear Square.

The research studies on machinery of training techniques in the field of farm crop production have been prepared by Mishra et al. [12]. Neural artificial (NN) networks used for the design of greenhouse gas (GHG) and output energy emissions for plant yield forecasting. In particular, ID3 and few optimisation algorithms, which are explained by artificial intelligence and machine learning algorithms, are used to develop a web-backed expert system with Java as its front and SQL as its backgrow to validate the crop symptoms.

Ganatra and Patel [1] have highlighted different stages of disease detection such as the acquisition of images, pre-processing, fragmentation of images, extraction of features and classification. In addition, pictures of the leaves were used to discuss various methods used for plant disease detection. In addition, the paper highlights some extraction and segmentation algorithms which can be used for the detection of plant disease.

Turkoglu and Hanbay [8] described the prevention of loss of productivity. The accurate diagnosis of plant diseases at the right time is essential in the prevention of losses and losses in productivity or reduced quantities of agricultural products. Methods based on machine learning can be used to resolve these problems. Deep learning, particularly used in image processing, offers many new applications in the field of precision farming. In this research we assessed the results using different approaches from nine powerful deeper neural network architectures for the detection of plant diseases.

Transfer learning and methods of profound extraction are used that adapt these models to the problem. In the present work for feature extraction and for further fine tuning, the used profound models are considered. The features gained using deep function extraction are then classified according to the methods supported by a vector machine (SVM), an extreme learning machine (ELM). The experiments are conducted with real disease and pest images data. In field of visual objects and machine learning, deep convolutional neural networks (CNNs) have brought significant gains. One of the greatest advantages improved by models and methods is that they are able to extract features without applying segmented methods.

3. PROPOSED SYSTEM ARCHITECTURE

Image processing is used to denoise the image and to extract the image characteristics. The ID3 algorithm collects the characteristics extracted from the sample seed data and locates the unwanted seeds. SVM then use soil data and properties of good seeds and predict the appropriate soil crop. Details of seeds along with growth, disease affecting them and preventive measures taken to overcome them are included in the samples seed data.

The system ultimately aims to predict crop yield based on whether and how data including geographical parameters are collected.

3.1 MODULE DESIGN

The working of the recommendation system incorporates the following modules such as:

- Image denoising
- Feature Extraction
- Seed classification
- SVM classifier

3.2 IMAGE DENOISING

Denoising is more important to image processing, analysis and applications than any other task. The aim of image denoising approaches is to preserve the details of an image and to remove the random noise as far as possible.

Image noise is only an unwanted image signal. The image of the seed is provided as input. Seeds include a variety of sounds like Gaussian noise, sound of salt and potatoes, speckle noise, sound of Poisson, wiener filter.

Gaussian noise is a statistically based ring with a probability density function (PDF) equivalent to the Gaussian distribution. That is the Gaussian-distributed values that the noise can take up. Salt/pepper noise is sometimes seen on pictures as a form of noise. The sound of an impulse is also known. Strong and sudden disturbances of the picture signal can result from this noise. It appears as white and black pixels that are slightly present.

Synthetic Aperture Ratio (SAR), active radar and optical coherence tomographic pictures, as well as the grainy salt and pepper pattern present in radar imagery. Noise from Poisson is a type of noise that is modelled by a process from Poisson. The discreet nature of the electrical charge causes shot sounds in electronics. Shot noise also occurs when photon counting occurs in optical equipment where the particle nature of light is associated with shot noise.

Several filters are available, including the Middle Filter, Mean Filter, Max Filter, Min Filter, Wiener Filter and Gaussian Filter.

The best filter for removing sounds from the image is among these filters. The Wiener filter is a filter used to produce a linear time-invariant (LTI) filter of an observed noisy process, assuming known stationary signal and sound spectrums, and additive noise, to estimate a desired or target random process. The wiener filter removes the noise and gives the denoted image when the seed picture is given.

3.2.1 Feature Extraction:

A feature is an information piece that is important to solve a certain application's computing task. Specific image structures such as points, corners or objects may be characteristic. The extraction of the feature describes the form data contained in the design so that a formal procedure makes the task of classifying the pattern easy. Each character in the feature extraction phase is represented by an identification feature vector. The main objective of extracting features is to extract a series of features that maximise recognition rates with the least number of elements and generate similar features for different instances of the same symbol. Many software packets for data analysis provide for extraction and reduction of features. Some simpler feature extraction techniques are available in common numerical programming environments, such as MATLAB, SciLab, NumPy and the language R. More specific algorithms are frequently provided in scripts or add-ons available to the public. There are also software packages for software apps specialising in extraction of features. The denoted picture is given as an input in this module, which extracts the characteristics of the image.

3.2.2 Seed Classification:

In this module, the samples and the characteristics extracted from the seed are inputs. The seed features are comparable to the sample seed data sets. The size, shape and colour parameters of the data sets are the ID3 algorithm. ID3 (Iterative Dichotomiser 3) is an algorithm for the generation of a decision tree from a dataset when it comes to decision-making. ID3 is the precursor of C4.5 and is typical in machine learning and the processing of natural languages. The ID3 algorithm starts as the root node with the original set S. It is used in all unused attributes of set S on each iteration of the algorithm to compute the H(S) entropy or the IG(S) gain of that attribute. The attribute with the smallest entropy value (or greater data gain) is then selected. The S set is then divided or partitioned to create subsets of data by the selected attribute. Every sub-set still uses the algorithm, taking into account only attributes not previously selected.

3.2.3 SVM Classifier:

A SVM is the classifier for discrimination formally defined by a hyperplane. In other words, the algorithm produces an optimal hyper-plane that categorises new examples with labelled training data (supervised learning). This hyperplanes is a line in two dimensional spaces that divides a plane into two parts where each class lays on both sides. SVMs can perform non-linear grading efficiently by mapping their inputs implicitly into high-dimensional spaces using what is called the kernel trick. An SVM model depicts the examples as spatially mapped items, which divide the examples of each category into the clearest possible divide. New cases are then mapped into the same area and predicted to be part of a category based on the side of the gap.

With the widest distance to the closest training data point of any class intuitively achieves the right separation. As the margin is generally bigger, the lower the classification error. In addition to performing linear classification, SVMs can effectively perform non-linear classification by mapping their entries in high-dimensional functional spaces using what is known by kernel tricks.

In this module, the sample characteristics are given as well as good seeds and soil data. The sample seed datasets contain the details of the seed and its growth under weather and soil conditions, respectively. The first is to compare the characteristics of sample seeds and good seeds to the fact that the soil is appropriate for plant production and to determine whether plant growth is good under conditions of climate and soil. The seeds then taken alone are suitable for soil. Secondly, seed characteristics are compared to sample data sets and disease predictions affect the crops, and preventive measures taken on a crop affected by illness also exist in the dataset.

Pseudo code for SVM classifier is given below:

Step 1: DS ←dataset of soil images with soil classes as labels

Step 2: SVM[*i*] ←one vs rest Support Vector Machine for class *i*

Step 3: for row in *D*

Step 4: SVM[row.class].train(row)

4. PERFORMANCE EVALUATION

The obtained results clearly show that the SVM performance with various filters, in both experiments. The Table.1 and Table.2 shows the obtained accuracy, in terms of PSNR, RMSE and MSE. The Table.3 shows the results of feature extracted from the dataset at training and testing phase. The Table.4 and Table.5 shows the results of accuracy of Training and testing set. The results shows that the proposed method achieves higher percentage of accuracy with average filter than median and wiener filter.

Table.1. Performance of various filters for Training Set

Image 1	PSNR	RMSE	MSE
Average filter	37.29	3.45	11.85
Median filter	42.29	1.96	3.80
Weiner filter	43.91	1.82	3.35

Table.2. Performance of various filters for Test Set

Image 1	PSNR	RMSE	MSE
Average filter	37.59	3.38	11.42
Median filter	40.29	2.96	5.50
Weiner filter	41.21	2.11	4.45

Table.3. Results of feature extracted from the dataset at training and testing phase

Label	Name	Train	Test
1	Coryneum beijerinckii (disease)	545	61
2	Apricot monilia laxa (disease)	76	9
3	Walnut leaf mite (pest)	62	7
4	Peach sphaerolecanium prunastri (pest)	310	35
5	Peach monilia laxa (disease)	278	31
6	Cherry myzus cerasi (disease)	175	19
7	Xanthomonasarboricola (disease)	129	14
8	Erwinia amylovora (pest)	193	21

Table.4. Accuracy of Training Set

Feature extracted	SVM - Weiner Filter	SVM - Median Filter	SVM - Average Filter
0	86.7	87.924	89.76
50	87.72	89.658	90.78
100	88.74	90.576	91.8
150	89.76	91.698	92.82
200	90.78	91.902	93.84
250	91.8	92.106	94.86
300	92.82	93.126	95.88

Table.5. Accuracy of Test Set

Feature extracted	SVM - Weiner Filter	SVM - Median Filter	SVM - Average Filter
0	83.2932	86.3532	88.74
50	84.4458	86.4858	89.76
100	85.0272	87.0672	90.78
150	86.6184	88.6584	91.8
200	86.8632	88.9032	92.82
250	88.3932	89.4132	93.84
300	89.5866	90.6066	94.86

5. CONCLUSION

In this paper, seed classification is carried out in order to classify the plant weeds among the seeds. Here the seed image and the sample seed datasets are entered. The seed image is given as an input in the preprocessing stage. An undesirable seed is eliminated by comparing the seed characteristics with the samples of seeds with the ID3 algorithm. The selection of appropriate cultivated soil is one of the reasons why crop production fails. As the data set of samples contains information about crop growth in soil, it helps to select the appropriate soil for seeds.

REFERENCES

[1] Nilay Ganatra and Atul Patel, "A Survey on Diseases Detection and Classification of Agriculture Products using Image Processing and Machine Learning", *International Journal of Computer Applications*, Vol. 180, No.13, pp. 1-13, 2018.

[2] Ashwini Rao, N. Abhishek Gowda and Rafega Beham, "Machine Learning in Soil Classification and Crop Detection", *International Journal for Scientific Research and Development*, Vol. 4, No. 1, pp. 1-12, 2017.

[3] David Camilo Corrales, Juan Carlos Corrales and Apolinar Figueroa-Casas, "Towards Detecting Crop Diseases and Pest by Supervised Learning", *Ingeniería Universidad Journal*, Vol. 19, No. 1, pp. 207-288, 2015.

[4] Narayan A. Joshi and Ketan Kotecha, "Plant Diseases Prediction Using Image Processing and Machine Learning Techniques: Survey", *International Journal of Computer Application*, Vol. 1, No. 8, pp. 1-12, 2018.

[5] E.M. Kamel, M. Wafy, H.M. Ibrahim and I.A. Badr, "Comparison of Different Classification Algorithms for Certain Weed Seeds Species and Wheat Grains Identification based on Morphological Parameters", *International Journal of Computer Science*, Vol. 12, No. 5, pp. 21-26, 2015.

[6] Ferhat Kurtulmus, Ilknur Alibas and Ismail Kavdir, "Classification of Pepper Seeds using Machine Vision based on Neural Network", *International Journal of Agricultural and Biological Engineering*, Vol. 9, No. 2, pp. 1-14, 2016.

[7] Godliver Owomugisha, Friedrich Melchert, Ernest Mwebaze, John A Quinn and Michael Biehl, "Machine Learning for Diagnosis of Disease in Plants using Spectral Data", *Proceedings of International Conference on Artificial Intelligence*, pp. 334-339, 2018.

[8] Muammer Turkoglu and Davut Hanbay, "Plant Disease and Pest Detection using Deep Learning-Based Features", *Turkish Journal of Electrical Engineering and Computer Sciences*, Vol. 27, pp. 1636-1651, 2019.

[9] Mukesh Chandra Arya and Ashish Semwal, "Comparison on Average, Median and Wiener Filter using Lung Images", *International Research Journal of Engineering and Technology*, Vol. 4, No. 2, pp. 1-15, 2017.

[10] P. Sachin D. Khirade, "Plant Disease Detection using Image Processing", *Proceedings of IEEE International Conference on Computing Communication Control and Automation*, pp. 1-6, 2015.

[11] Shivnath Ghosh and Santanu Koley, "Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks", *International Journal on Recent and Innovation Trends in Computing and Communication*, Vol. 2, No. 2, pp. 45-52, 2015.

[12] Subhadra Mishra, Debahuti Mishra1 and Gour Hari Santra, "Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper", *Indian Journal of Science and Technology*, Vol. 9, No. 38, pp. 1-13, 2016.

[13] Tekalign Tujo, Dileep Kumar, Elifenesh Yitagesu and Meseret Girma, "A Predictive Model to Predict Seed Classes using Machine Learning", *International Journal of Engineering Research and Technology*, Vol. 6, No. 8, pp. 22-28, 2017.

[14] Vishal Vats, Arun Kumar and Naveen Kumar, "Efficient Crop Yield Prediction using Machine Learning Algorithms", *International Research Journal of Engineering and Technology*, Vol. 5, No. 6, pp. 1-16, 2018.

[15] Yun Hwan Kima, Seong Joon Yooa, Yeong Hyeon Gua, Jin Hee Limb, Dongil Hana and Sung Wook Bai, "Crop Pests Prediction Method using Regression and Machine Learning Technology: Survey", *IERI Procedia*, Vol. 6, pp. 52-56, 2018.