Leaf Disease Detection of SoyBean Plant using Machine Learning Algorithms

M.Sowmiya, C. Thilagavathi

Abstract*: *Nature is an important phenomenon of this physical world. This nature includes plants, trees, animal, humans and several other organisms. Amongst them plants are the most important organisms in the world. They are specialist of creating their food by themselves and also they are the notable components in food chain. They also serve the nature and its organisms in a tremendous way. Hence it is necessary to protect our nature in an efficient manner so as to maintain the food chain. Our technology development has given much advancement in the field of agriculture. This paper deals with the analysis of various machine learning algorithms, by applying the algorithms on the plant data set. Sample sizes of collected data set are used to train the algorithm and the results are evaluated accordingly to estimate the better implementation of machine learning algorithm.

Keywords* : *Machine Learning, Leaf Disease Detection, SVM

I. INTRODUCTION

Agriculture is the backbone of Indian republic. The economy of the nation lies in crop production. The increasing demands on food and various crops have made agriculture boon tremendously in the nation. The advancements in technology have started helping the growth of agriculturalists in an efficient manner. Some kind of expert system is needed to monitor the plant growth. Machine learning is a technology that makes the computer systems learn data and acquire knowledge out of it and thus enabling them to learn by themselves.

Many machine learning algorithms exists to make the computers smarter. Sample data can be trained into the machines to understand the knowledge of the data. This paper is implemented by considering the few of the techniques of machine learning. Machine learning algorithms are generally of supervised and unsupervised knowledge algorithms. The proposed paper considers supervised learning techniques by training the algorithms with 50 sample data sets of soybean plant.

Several procedures such as logistic regression, linear discriminant analysis, support vector machines, naïve bayes, random forest are available in supervised learning. The histogram of the plant has been trained to the algorithm and the machine obtains knowledge from it. Along with that, certain attributes of soybean is also trained to the computer system, thus helping the algorithms to identify between a diseased leaf and a healthy leaf. Soybean is a species of legume and it's a native crop of East Asia. It is an annual legume of pea family. Several diseases, along with soil-borne water mould and shoot rot, pod and stem disease, frogeye leaf spot, brown spot, furry decay, Cercopsora leaf blight and crimson seed stain, and Sclerotinia stem rot (white mold) will have an effect on the growth of the plant. Hence the proposed system enables the user to identify the easier way of early detection of diseases in soybean plant and also to recover from it by applying proper agricultural methods.

II. LITERATURE REVIEW

Anuradha Badage has proposed a model for disease detection of wheat plant. The work was carried out into two phases namely, training phase and monitoring phase. The monitoring phase detects the disease of wheat plant using the Canny's Edge detection algorithm. Based on the detection of disease the effective decision support system intimates the farmer to take necessary action.

T. Rumpf, A.-K. Mahleinb, U. Steiner, E.-C. Oerke, H.-W. Dehne, L. Plümer has introduced SVM algorithm to detect the disease in sugar beet leaves. The algorithm trains with the given data set and showed 97% accuracy in disease detection.

Shima Ramesh Manniath, RamachandraHebbar, Niveditha M, Pooja R, Prasad Bhat N, Shashank N, Mr. P V Vinod presented training of large data sets of papaya leaf. The histogram plots of a healthy leaf and an infected leaf is trained into the algorithm and the incoming images of leaf are investigated using various machine learning algorithms. Among the executed algorithms, it was evident that Random Forest algorithm shows 70% accuracy in early detection.

K. Indumathi, R. Hemalatha,S. AashaNandhiniand S. Radha, "Intelligent Plant Disease Detection SystemUsing Wireless MultimediaSensor Networks" has developed a disease detection system for spotting and categorizing the disease in plants. The received images are transmitted through nodes and the categorization is done using support vector machine classifier. Then, results are compared with different existing algorithms and possess 98% accuracy. This technique was even analyzed in real time.

DavoudAshourloo, Hossein Aghighi, Ali Akbar Matkan, Mohammad Reza Mobasheri, and Amir Moeini Rad has made a comparison on three distinctive regression strategies of dev ice learning. This paper detects the leaf disease detection in wheat rust crop. The paper taken into consideration 25 sample information units of wheat rust crop.

Revised Manuscript Received on January 5, 2020

***** Correspondence Author

Mrs.M.Sowmiya*, Assistant Professor, Department of IT, M.KUMARASAMY COLLEGE OF ENGINEERING, KARUR, Mail id.: sowmiyam.it@mkce.ac.in.

Mrs.C.Thilagavathi, Assistant Professor, Department of IT, M.KUMARASAMY COLLEGE OF ENGINEERING, KARUR, Mail id.: thilagavathic.it@mkce.ac.in .

III. PROPOSED WORK

The proposed work attempts to monitor the plant leaf disease detection of Soybean plant. A small subset of SoyBean is considered for sample data set. Soybean is an annual crop, initiated from Northeast China. It has woody stems and the leaves are arranged alternatively in it. Soybean leaves are vulnerable to many bacterial and fungus disease. Some of which include bacterial blight, bacterial pustule, anthracnose, Cercospora Leaf Blight and purple Seed Stain etc., Our system is designed in a way to detect the leaf disease in plants using various classification algorithms.

Machine learning algorithms play a vital role in agricultural field. Its effective and efficient techniques have made the agricultural farming as smart farming. Artificial intelligence has made its introduction in developing robots for agricultural purposes. Nowadays, robots are used for seed sowing, water feeding and several other agricultural activities. Machine learning also monitors the growth of plants for early detection of plant diseases. We have several machine learning algorithms such as classification, clustering etc., This paper makes data set to be executed on classification algorithms. The system begins by first collecting the sample data sets, processing them by clearing all the missing values. The flow continues by feeding the data set into classification algorithms and finding the accuracy of each algorithm. Each of the classification algorithms are executed on the trained data sets. Some of the categorization procedures such as Decision Trees, Naive Bayes Classifiers, SVM are considered to apply. Here we study the implementation of algorithm using R.

Decision Tree Algorithm:

This algorithm is one among the supervised learning procedures. It is applicable for both classification and regression algorithms. It has the capability to accept categorical data values as well as continuous data values. Categorical data sets involve taking decisions like yes/no. Continuous variables will have target variable. This algorithm generates a tree structure to make decisions on whether a leaf is infected or not. The decision tree consists of nodes and links connected to the nodes. Each node represents a predictor variable and each link represents a decision made on the predictor variable. Similarly leaf node indicates the outcome. Some of the important terminologies exist in decision tree such as:

Root node: The whole data set will be divided into two / more data sets.

Splitting: A node will be divided into multiple sub nodes.

Decision node: A subnode further splitting into a sub node

Leaf node: Nodes do not split decision tree algorithms are easy to understand and develop.it is very useful in data exploration. The algorithm is trained with a data set.

```
library(rpart)
 x <- cbind(data1,data2)
# developing tree
f <-rpart(data2~., data = x, method="class")
summary(f)
#Finding Output
found= predict (f, x_test)
# Constructing the classification tree with rpart
t<-rpart(class~.,data=new,parms= list(loss = penalty.matrix),method = "class")
Visualization:
rpart.plot(t, nn=TRUE)
```

Naïve Bayes Classifiers:

It is a classification algorithm that is constructed with help of Bayes theorem. It follows a probalistic approach. It follows a concept that the predictor variables are not dependent to each other. It obtains an output of either true or false. In the proposed system we have choosen Soybean data set with 47 instances and 35 attributes. Let us consider the data to predict whether the plant is diseased or not. To begin with, we have to load the data set into R.

```
d<-read.csv("D:\19-0\RESEARCH\soybean-smalldata.csv")
```

then set the output variable

```
d$Outcome<-factor(d$Outcome, levels = c(0,1), labels = c("False", "True"))
#Learning the assembly of the facts
str(d)
```

Since our chosen information set do not Have missing values data cleaning process can be skipped.

Now we can visualize the data based on exploration.

```
ggplot(d, aes(leaves, colour = Outcome)) +geom_freqpoly(binwidth=1)+ labs(title="Divergence of leaves through Result")
```

Similar conception can be done on all other attributes related to this study. The subsequent step is to load the e1071 package deal that holds the Naive Bayes characteristic. This is an in-built feature furnished by means of R. The final output shows that we constructed a Naive Bayes classifier which could predict whether or not a leaf is diseased or not, with an accuracy of about 73%.

```
#Design Adjustable routine
A <- varImp(model)
plot(A)
```

Support Vector Machines:

SVM is one among the supervised algorithms. It is specifically used to categorize the data into one of a kind class. SVM creates a hyperplane and acts as a choice boundary between the various classes. SVM can be used to generate multiple setting apart hyperplanes such that the statistics is divided into segments and each phase contains only one type of data. We begin of by drawing a random hyperplane and then you check the distance among the hyperplane and the closest records factors from every class. These closest information factors to the hyperplane are called guide vectors. If data is of non linear form we, transform the 2D data into 3D space. To begin with, we install caret package which is very helpful in data processing steps. It has all tools needed for data preprocessing.

```
install.packages("caret")
```

Now we add it into our library

```
library('caret')
```

Then we intend to add the data set, which is more similar to same way as previous algorithms. Then we perform all necessary steps and finally test our model. Our correctness on the check set is 87.78 %.

IV. RESULTS AND DISCUSSION

The problem statement in the proposed system is to predict whether the leaf is diseased or not. For this purpose, we chosen the best, popularly used classification algorithms such as Decision Tree Algorithm, Naïve Bayes Classification Algorithm and Support Vector Machine algorithm. All these algorithms are implemented using R. the data set has been trained and the corresponding models are created for further testing of data set. Finally, the models are tested with test data set. Among the tested algorithms, it is evident that SVM shows high performance of 87.78% in predicting the diseased leaf. Fig 4.1 shows the comparative results of various machine learning algorithms for our proposed work.

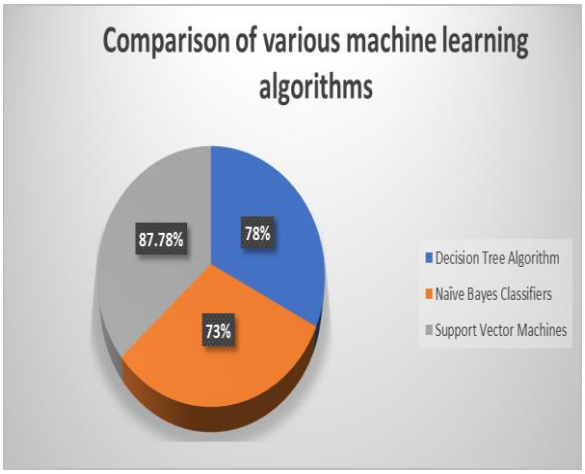

Fig.1 Comparitive results

V. CONCLUSION

The objective of this paper is to recognize the abnormalities inside the leaf statistics set to predict its nature of boom in their greenhouses or herbal environment. The statistics set and different attribute data are fetched from UCI Machine studying repository. Here we considered only few attributes to predict the nature of the leaf, we can even include some additional attributes if the data set to ensure high accuracy. It is also evident that SVM & Decision tree algorithm shows good performance results when compared to Naïve Bayes Algorithm. The future research can include some more attributes of the data set to ensure high accuracy rates.

REFERENCES

1. Shima Ramesh, Mr. Ramachandra Hebbar, Niveditha M, Pooja R, Prasad Bhat N, Shashank N, Mr. P V Vinod, " Plant Disease Detection Using Machine Learning",DOI10.1109/ICDI3C.2018.00017, 2018.
2. Anuradha Badage, " Crop Disease Detection using Machine Learning : Indian Agriculture",2018.
3. T.Rumpf., A.K.Mahlein, U.Steiner, E.C.Oerke, H.W.Dehne, L.Plumer, " Early Detection and Classification of Plant Disease with Support Vector Machines based on Hyperspectral Reflectance", 2010.
4. K. Indumathi, R. Hemalatha, S. Aasha Nandhini and S. Radha,, " Intelligent Plant Disease Detection System Using Wireless Multimedia Sensor Networks",2017.
5. Davoud Ashourloo, Hossein Aghighi, Ali Akbar Matkan, Mohammad Reza Mobasheri, and Amir Moeini Rad, " An Investigation Into Machine Learning Regression Techniques for the Leaf Rust Disease Detection Using Hyperspectral Measurement",2016.
6. S. S. Sannakki and V. S. Rajpurohit," Classification of Pomegranate Diseases Based on Back Propagation Neural Network," International Research Journal of Engineering.