Journal of Green Engineering (JGE)

Volume-10, Issue-12, December 2020

Classification and Prediction in Agricultural Domain Using the Linear Discriminant Analysis (LDA) Algorithm

[1]M.C.S. Geetha and [2]I. Elizabeth Shanthi

[1]Assistant Professor, Department Master of Computer Applications, Kumaraguru College of Technology, Coimbatore, India.
E-mail: research.mcsgeetha@gmail.com
[2]Associate Professor, Department Computer Science, Avinashilingam University for Women, Coimbatore, India.
E-mail: shanthianto@gmail.com

Abstract

At present, computers have been in agriculture for the automation of different applications. The expert and decision support system is used for taking the critical decision about crop yield and measures for plant protection. It is a tedious process for farmers to predict the cultivated area manually when it is enormous in acres. The proposed scheme presents a result of frequently watching the cultivated region and offers automatic disease discovery using data mining techniques. For example, data mining techniques help to detect or forecast crops diseases, manufacture, and loss. The disease patterns are used to recognize diseases. Generally, the classification rules and relationships are used for acquiring knowledge from empirical data of diverse dataset. This paper explores what Classification rule and prediction can do in the agricultural domain. This paper goal to detect classification rules for the Indian Banana, Paddy and Sugarcane using the Linear Discriminant Analysis (LDA) algorithm. It is compared with standard machine learning and neural network algorithms for example artificial neural network, Support vector machine and Decision Tree.

Keywords: Data Mining, Paddy Crop, Diseases, Classification, Expert System, Agriculture

Journal of Green Engineering, Vol. 10_12, 13374-13392.

1 Introduction

Agriculture being the backbone of economic improvement of the country, the farmers need to monitor the crop for optimum yield per year periodically. Agricultural production, its quality and quantity are greatly affected by various crop diseases [1].

The information about new varieties of the crop is very much useful to farmers for the decision-making process in adopting the varieties that will improve production. Sometimes, the new types can use the farmer's resource outlays as such without any change. The information is essential to both farmers and seed sellers and must be precise. Data mining is the taking out of useful knowledge from unknown data of enormous databases. This helps to predict future behaviour to take proactive knowledge-based decisions [2].

Linear Regression model produces output for different information and tests the model based on mathematical relationship intrinsic to previous dataset history. This model predicts the yield despite the limitation of the agriculture field with sparse and incomplete data [3].

Various technologies, tools and algorithms strengthen the agricultural study, which is very important for the increase in economic income. One such technique is data mining which is used to analyze, evaluate and derive patterns and knowledge from predicting the findings. The prediction results are used for further research to improve production [4].

The term *Knowledge Discovery in Databases (*KDD) is utilized to take out knowledge from huge databases and apply in data mining methods. Many researchers have applied machine learning patterns, statistics and pattern recognition to solve the problems in agriculture. The data can be of symbolic or numeric and also both. The machine learning schemes can derive relationships between the data and statistically analyzed to bring out its significance [5].

The data mining techniques are of predictive and descriptive where again the predictive type is distinguished into classification, regression and time series. These techniques highly support the prediction of yield and crop diseases [6]. It is also used to study crop and soil management. This paper describes four machine learning algorithms Bayesian Network, Neural Network, MLR (Multiple Linear Regression), and SVM (Support Vector Machine) used for various cases. Based on the forecast and prediction, disease control measures are taken to avoid loss in crop production [7].

The disease is predicted based on the symptoms and reactions shown by the systems. This automatic disease prediction and control systems are of significant concern for the scientists for improvement in agriculture [8]

The remaining of paper is structured as follows: Related Work provides in Section 2, the Materials and Methods provides in Section 3, Section 4 provides the proposed method for crop yield forecast,

Section 5 explains the Result & Discussion of three crops and Section 6 presents the Conclusion of the work.

Machine learning is defined as learning process from the environment and study of a system without any specific program which is then used to improve system performance. The learning methods are of following types

1) Supervised learning: The learning process uses prior knowledge as input and produces the output pattern. This is called self-supervised learning.

2) Unsupervised learning: There is no prior knowledge to check the output patterns and called an unsupervised study.

3) Reinforced learning: It is of both supervised and unsupervised learning where knowledge is used for checking whether the output is correct or incorrect [9].

2 Related Work

Cora B. Perez-Ariza et al. [10] uses Bayesian Network for the prediction of disease on coffee farming so that measures can be taken for improving the yield. In [11], classification is used to classify disease in leaf. The k-NN algorithm is utilized for forecasting the noisy input and used to predict the classes.

In [12], the system uses an image process for studying the disease of two plants, grapes and wheat. Backpropagation (BP) networks are used to classify diseases. On using principal component analysis (PCA), the expectation exactness and fitting precision are achieved as 100%. The same study is done on decreasing the principal component of the element where the forecast precision is 100% for grapes and wheat, and expectation exactness is 97.14% for grapes and 100% for wheat.

In [13], the author used artificial neural networks (ANNs) to extract patterns from multivariate data with complex relationships and without any prior assumptions. The author in [14] used Adaboost & RF and enhanced the precision of feeble classifiers. Similarly, at [15], the author utilized NN, SVM, and DT classifiers and achieved better prediction accuracy than KNN and GNB.

Uno et al. [16] used an ANN for a sample of the corn field. The crop was studied for different weed control methods under different amount of nitrogen usage. The author compared the ANN system with step-wise multiple linear regression (SMLR), and the prediction root mean squared error rate obtained as 20%.

El-Telbany and Warda [17] applied the decision tree using C4.5 algorithm for classifying the diseases affecting Egypt rice variety. They used seven attribute data for collecting the dataset. The attributes are selected based on entropy and gain. The C4.5 is compare with ANN algorithm also the precision attained is 97.25% and 96.4%, respectively

The author Diriba and Borena [18] used the classification algorithm for predicting the crop production in Ethiopia. The dataset was obtained from the research department of Ethiopia with eleven agronomic attributes.

The amount of fertilizer is found to influence the crop yield. The study was done using classification algorithms as j48, REPTree, and random forest. The prediction accuracy obtained was 83%.

Ashwinirani et al. [19] studied the yield of sugarcane using C4.5 and PCA classifier. The yield is predicted using weather data along with some basic parameters of cultivation.

Veenahadri et al. [20] used a C4.5 for the study of the soya bean yield in Bhopal district with parameters like environmental factors such as humidity, rainfall and temperature are considered rather than agricultural factors. The decision tree for high and low yield is obtained. On using K-means 90% prediction accuracy and on using kNN classifiers, 76% prediction accuracy was collected.

3 Materials and Methods

Classification is one of the data mining techniques to group or classify the data into predefined label or taxonomies. It is used to predict the class under which a data record falls correctly. In the agricultural field, it is mostly used to predict crop yield or loss. The following are the introduction to various classification algorithms.

Decision Tree (DT) classifier is a tree structure form. The tree nodes are fixed as attributes which are linked to from the tree. The links are used to set split points using entropy and information gain of the attributes calculated for each node. The nodes are repeatedly split at the end of which each node belongs to the single class label [21].

C4.5 [22] is a classification algorithm by Ross Quinlan as an expansion of Quinlan's previous ID3 algorithm. The algorithm traverses top-down induction of the tree. The user provides the training set based on which the testing set is classified. At each node, the attribute is tested when root specifies the attribute to be started with for classification. The leaves decide the final classification classes. The attribute to be chosen for the test is decided based on an information-theoretic heuristic test that minimizes the entropy. The proportion of the positive (p) and negative cases (q) at each node is calculated using the formula

$$-p\log_2 p - q\log_2 q$$

C4.5 can create tree effectively with more predicting power with a prediction error rate of 1.5% on test data

Support vector machines (SVM) are a statistical method with strong hypothetical base and high optimization capacity. The technique enhances generalization capacity despite various issues of machine learning model such as local minimization, dimensionality problem, over fitting and non-linear points. The SVM classifies the feature vectors which are expressed in the plane that separates the classes as per the plane maps.

The SVM is advantageous than other methods in handling high dimensional dataset with robustness and flexible in fitting the feature vectors with non-linear relationships [23].

The neural network is a classifier that maps the information into the correct class label by learning the weights of the feature in feature space. The weight trained is utilized to forecast the class of experiment instance. It works on three-layers such as input, hidden and output layer. The amount of input vector patterns represents the amount of neuron at the input layer. The hidden layer uses a sigmoid function for the process, and neurons at an output layer correspond to output class labels [24]. In [25,27] the authors discussed about discovering malevolent URLs based on binary classification using ada boost algorithm and also they have analyzed the digital DNA sequencing engine for ransomware discovery based on ML. The analysis of diabetic retinopathy based on multi-level set segmentation algorithm with feature extraction based on SVM with choosy features [26]. Choice of best hyper-parameter values of SVM for sentiment analysis utilizing nature-inspired enhancement techniques was discussed in [28]. A proficient apriori algorithm for recurrent pattern mining based on mapreduce at healthcare data was proposed in [29-30].

4 Proposed Linear Discriminant Analysis for Crop Prediction

In this research, the classification step for predicting the classes is performed by using decision tree induction algorithm, which is a top-down approach. The nodes hold splitting criteria, and edges represent test outcomes. The leaves are the classification results with class labels. In the proposed system, based on crop agronomic and meteorological parameter, the three-class labels are low, medium and high yield.

Discriminant analysis is a classification technique used to analyse the research data when two or more group or cluster is known based on which further group is classified. The method is suitable when the reliant variable is categorical; also the predictor variable or autonomous variable is an interval type.

Discriminant analysis is similar to regression analysis which finds the relation between the predictor and dependent variable with which the predictor variable value is calculated. Logistic regression is a two-class classification problem that works like the least square regression. The discriminant analysis creates a maximum difference between the groups. This is extended as a Linear Discriminant Analysis (LDA) for the multi-class classification problem.

4.1 Representation of LDA Models

LDA assumes that data is Gaussian type, which when the variable is

single, there is mean and variance for each class. But when there is multiple variable, each class will have means and covariance matrix. Each variable form the bell shaped curve while plotted.

Each variable has some variance. The value of the variable varies by the same average from the mean. LDA estimates mean and variance for data of each class. The mean (m) of each input (x) for each class (y) is given by dividing the sum of values by the total number of values as follows

$$my = 1/ny * sum(x)$$

where ny is the number of samples with class y.

The variance is given as an average of squared difference of each value and mean. It's given as follows

$$\sum^2 = 1/(n-y) * sum((x-m)^2)$$

Where $\sum^2$ is variance across all inputs x, n is the number of instances, y is the number of classes and m is mean of x.

4.2 Creating Forecast with LDA

LDA predicts the class of new inputs by calculating the probability of belonging to each class. The maximum probability is given as the output class. The probability is evaluated using Baye's theorem as followed. The probability that a data belong to the class (y) for the given input (x) is given as

$$P(Y=x|X=x) = (Ply * fy(x) / sum(P|| * f(x))$$

Where Ply is the base probability of class (y) observed from training data. This is the prior probability and is given as

$$Ply = ny/n$$

The f(x) is the estimated probability of x belonging to a class and is an estimate using Gaussian distribution function. The discriminant function as classification output is given as

$$Dy(x) = x * (my/\sum^2) - (my^2/(2 * \sum^2)) + ln(Ply)$$

Dy(x) is the value for the class y when x is given input, my, Ply and $\sum^2$ are estimated from the provided data.

The calculation of discriminant function is involved in the stepwise algorithm of LDA shown below

4.3 Steps of LDA Algorithm

Step 1: Input the sample or training data and test data
Step 2: Find prior probabilitie pi of expected class with population πi
Step 3: Linear discriminant analysis is done for homogenous variance-covariance matrices
Step 4: Find the conditional probability density functions f(X|πi) .
Step 5: Compute discriminant functions for classifying the input to known class
Step 6: Classify test observations with the group or community

The algorithm is explained as follows

Step 1: Training data are grouped with known group memberships.

Step 2: The prior probability Pi gives the expected number of class members with population πi with three choices equal prior, arbitrary prior for which p1 +p2+p3+….+pn =1 for n instances and estimated prior given as pi = ni * N where ni is number of observation and N is the total number of observation

Step 3: Using Bartlett's test variance-covariance matrices are found for applying

Step 4: With the following four assumptions that

i) Data from group i will have a common mean vector,

ii) Data from a group will have a common variance-covariance matrix,

iii) Each subject are sampled independently and

iv) The data are multivariate with normal distribution, the conditional probability density function is estimated

Step 5: Discriminate function is estimated to classify the new data into one of the known class

Step 6: Classify the test data with the community it falls

5 Results and Discussion

The training and testing data of 100 instances each for the banana, paddy and sugarcane are taken as input dataset, and the proposed linear discriminant analysis algorithm is compared with a ANN, C4.5 decision tree, and SVM techniques. The results and the description of the Banana are explained here. Initially, Six attributes named as Q1, Q2, Q3, Q4, Q5 and Q6 are considered. Each attribute represents five symptoms. The illustration is shown below for the Q1

1) Yellowing of a lowermost leaves opening from edge to midrib of the leaves

2) The yellowing enlarges upwards, also ultimately, heartleaf unaccompanied leftovers green for for a moment, also it is affect.

3) The leaves smash close to the base also dangle approximately pseudostem.

4) Longitudinal dividing of pseudostem.

5) Premature symptom appears on a third leaf from the top.

The sample Disease Names taken for the research is given below

- Anthracnose
- Banana_bract_mosaic_virus
- Banana_streak_disease
- Bunchy_top/curly_top
- Infectious_chlorosis

- Moko_disease/bacterial_wilt
- Mycosphaerella_leaf_spot_sigatoka
- Panama_wilt
- Tip_over_or_bacterial_soft_rot
- Moko_disease/bacterial_wilt

The Sample of 100 instances is taken as Training data for the banana. The below-mentioned values are taken as Q1, Q2, Q3,Q4,Q5, Q6, and the disease name are shown in Table 1.

Table 1 Attributes with the Disease Name

Attributes						Disease Name
Q1	Q2	Q3	Q4	Q5	Q6	
2	5	4	1	3	5	Panama_wilt
3	2	4	1	5	2	Banana_streak_disease
5	1	5	1	1	3	Anthracnose
1	2	2	5	2	3	Moko_disease/bacterial_wilt
3	4	3	5	4	4	Anthracnose
2	5	2	1	4	4	Tip_over_or_bacterial_soft_rot
3	1	1	5	1	1	Infectious_chlorosis
1	2	4	5	4	4	Anthracnose
5	4	3	3	4	3	Anthracnose
1	4	5	3	5	4	Banana_bract_mosaic_virus
4	4	1	2	2	3	Infectious_chlorosis
3	1	4	3	1	1	Mycosphaerella_leaf_spot_sigatoka
4	2	1	3	2	2	Tip_over_or_bacterial_soft_rot
3	2	4	3	2	3	Banana_bract_mosaic_virus
2	2	2	3	3	3	Infectious_chlorosis

The Table-2 depicts the Banana Training and Testing Error measures.It is compared with C4.5, SVM,ANN and DA algorithms

Table 2 Testing Error Measures for Banana

Algorithms	C4.5	SVM	ANN	DA
Noof instances	100	100	100	100
Correctly classified	54	78	56	100
Incorrectly classified	46	22	44	0
Kappa Statistic	0.4745	0.7506	.5017	1
Mean absolute error	0.1194	0.1752	0.1221	0

Root mean squared error	0.2443	0.2844	0.243	0
Relative absolute error	60.8323 %	89.279 8 %	62.2053 %	0.000 4%
Relative Squared error	78.0154 %	90.818 2 %	77.5849 %	0.000 9%
Coverage of cases	100 %	100 %	84%	100%
Mean rel. region size (0.95 level)	31.2222 %	84.111 1%	43.6667 %	11.11 11%

The proposed DA algorithm is compared with C4.5, SVM, ANN and the performance metrics are compared with the existing algorithms. The comparision is given in the Table-3

Table 3 Comparitive Measures of Each Technique for Banana

Algorithms	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
C4.5	0.54	0.065	0.592	0.54	0.511	0.931
SVM	0.78	0.029	0.792	0.78	0.777	0.967
ANN	0.56	0.055	0.631	0.56	0.551	0.875
DA	1	0	1	1	1	1

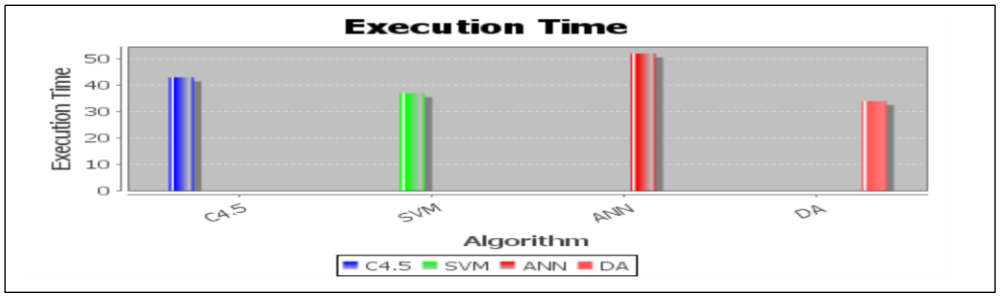

Figure 1 Execution Time Comparison

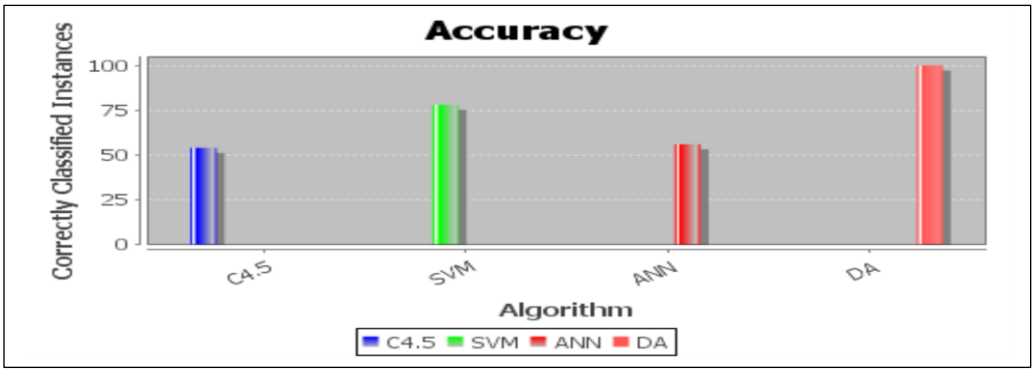

Figure 2 Accuracy Comparison

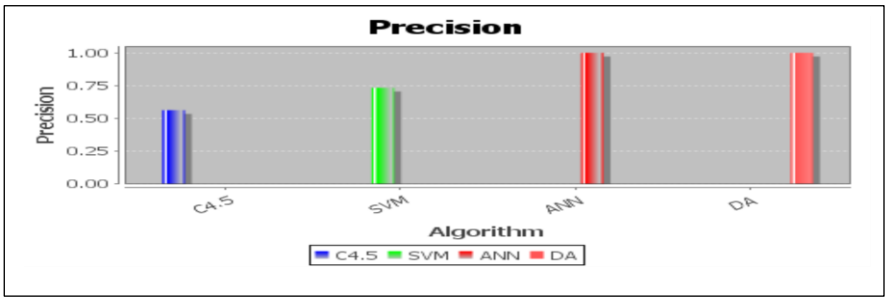

Figure 3 Comparison of Precision

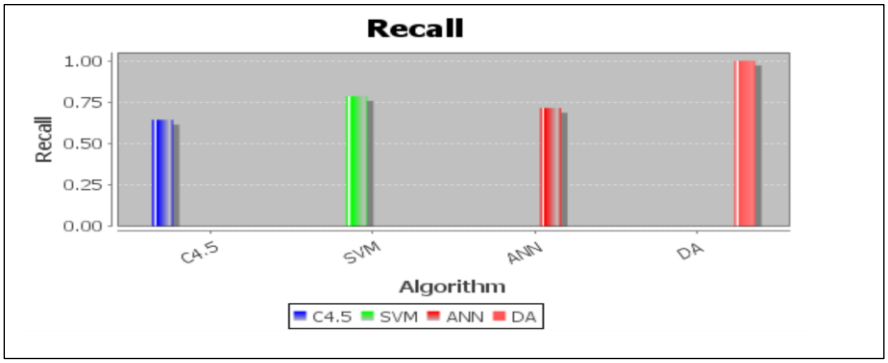

Figure 4 Comparison of Recall

The results are visually shown through the figures 1,2,3 and 4.In Figure 1, the execution time of the proposed algorithm is compared with the existing algorithms. As well as in Figure 2, 3 and 4, the accuracy, precision and recall are compared. Through these performance metrics, it is proven that the proposed algorithms are betted then the existing algorithms in terms of performance.

As like banana, we have worked for Paddy crop also. The results and the description of the paddy are explained here. Initially, Nine attributes are taken as Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 and Q9. Each attribute represents five symptoms. The illustration is shown below for the Q1

1) The disease can infect paddy at all growth stages and all aerial parts of the plant (Leaf, neck and node).
2) Among the three leaves and neck, infections are more severe.
3) Small specks originate on leaves - subsequently enlarge into spindle-shaped spots(0.5 to 1.5cm length, 0.3 to 0.5cm width) with the ashy center.
4) Several spots coalesce -> big irregular patches
5) Seedling wilt or kretek

The sample Disease Names taken for the research is given below

- Bacterial Leaf Blight - Xanthomonas oryzae pv.oryzae
- Blast - Pyricularia grisea (P. oryzae)
- Brown Spot - Helminthosporium oryzae

- False Smut - Ustilaginoidea virens
- Grain discolouration - fungal complex
- Leaf streak - Xanthomonas oryzae pv. Oryzicola
- Rice tungro disease - Rice tungro virus (RTSV, RTBV)
- Sheath Blight - Rhizoctonia Solani
- Sheath Rot - Sarocladium oryzae

The Sample of 100 instances is taken as Training data for the paddy. The below-mentioned values are taken as Q1, Q2, Q3, Q4,Q5, Q6,Q7,Q8, Q9 and the disease name are shown in Table 4.

Table 4 Attributes with the Disease Name for Paddy

Attributes									Disease
Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	
1	1	3	1	2	5	2	5	1	Blast
1	2	3	2	1	1	2	4	1	Brown_Spot
2	5	5	3	3	1	4	5	2	Grain_discolouration
5	3	4	1	4	4	5	5	1	Rice_tungro_disease
1	5	1	1	4	1	1	2	3	Bacterial_Leaf_Blight
1	5	2	2	1	2	5	1	1	Blast
3	3	2	1	2	3	2	4	2	Grain_discolouration
1	3	1	5	3	4	4	2	2	Sheath_Blight
3	1	3	4	4	3	1	1	4	Grain_discolouration

The Table-5 depicts the Paddy Training and Testing Error measures. It is compared with C4.5, SVM, ANN and DA algorithms

Table 5 Testing Error Measures for Paddy

Algorithms	C4.5	SVM	ANN	DA
No of instances	100	100	100	100
Correctly classified	51	98	72	100
Incorrectly classified	49	2	28	0
Kappa Statistic	0.4429	0.9774	0.6837	1
Mean absolute error	0.1283	0.173	0.0958	0
Root mean squared error	0.2532	0.2806	0.2043	0
Relative absolute error	65.1564 %	87.8607%	48.6444 %	0.0001%
Relative Squared error	80.7308 %	89.4615%	65.1429 %	0.0001%
Coverage of cases	100%	100%	89%	100%
Mean rel. region size (0.95 level)	40%	84.1111%	38.4444 %	11.1111%

The proposed DA algorithm is compared with C4.5, SVM, ANN and the performance metrics are compared with the existing algorithms. The comparison is given in the Table-6

Table 6 Comparative Measures of Each Technique for Paddy

Algorithms	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
C4.5	0.51	0.068	0.536	0.51	0.467	0.908
SVM	0.98	0.003	0.983	0.98	0.98	0.998
ANN	0.72	0.032	0.751	0.72	0.713	0.901
DA	1	0	1	1	1	1

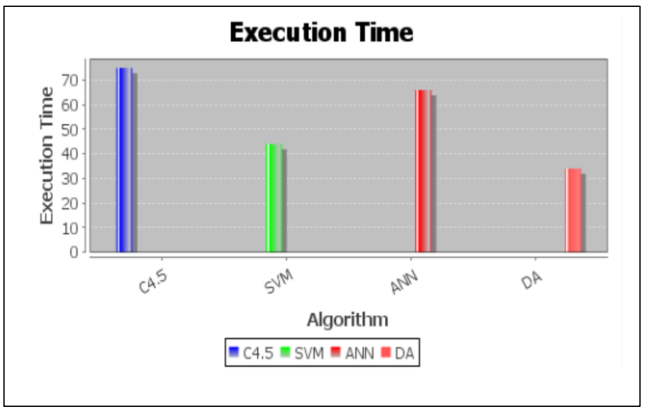

Figure 5 Comparison of Execution Time

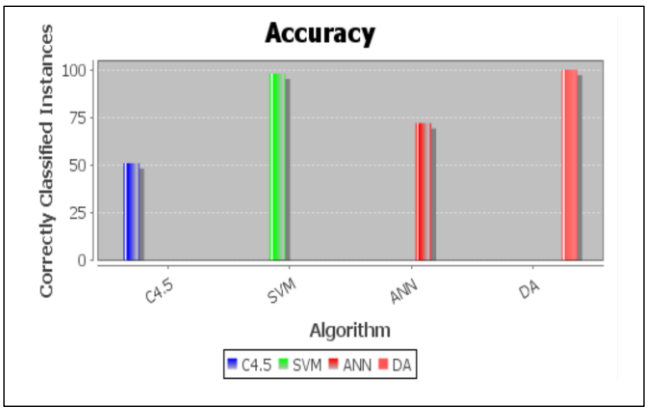

Figure 6 Comparison of Accuracy

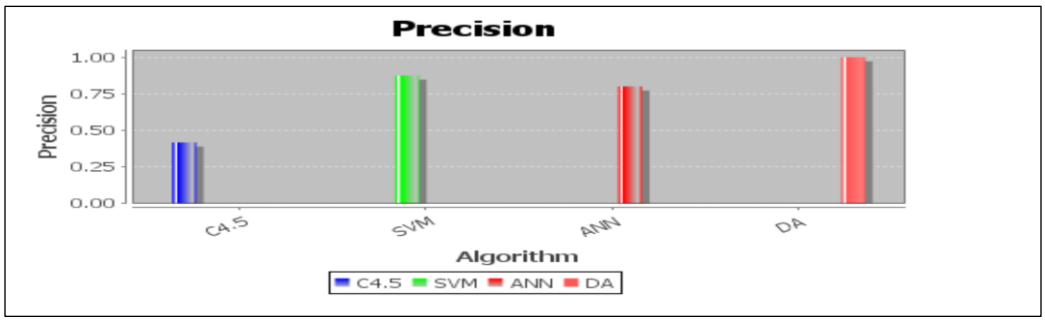

Figure 7 Comparison of Precision

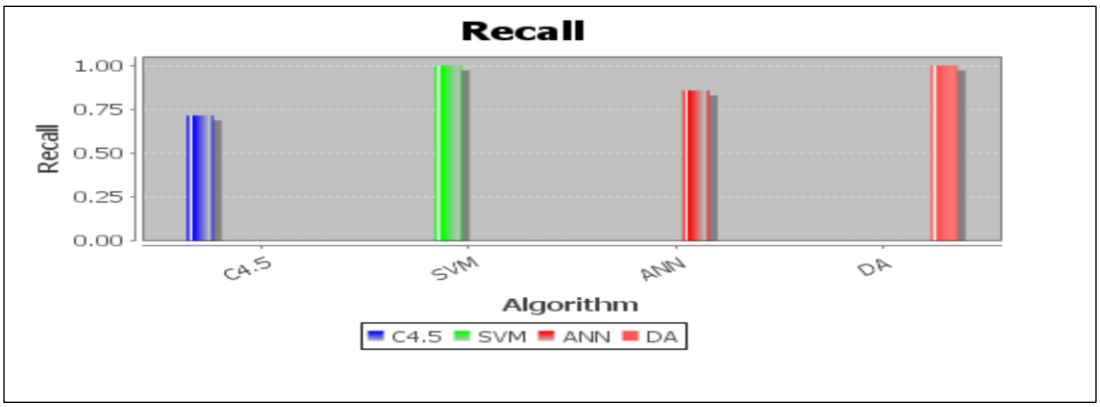

Figure 8 Comparison of Recall

The results are visually shown through the figures 5,6,7 and 8. In Figure 5, the execution time of the proposed algorithm is compared with the existing algorithms. As well as in Figure 6, 7and 8, the accuracy, precision and recall are compared. Through these performance metrics, it is proven that the proposed algorithms are betted then the existing algorithms in terms of performance.

We have experimented the same algorithm with the sugarcane crop also. The results and the description of the sugarcane are explained here. Initially, Seven attributes are taken as Q1 Q2 Q3 Q4 Q5 Q6 and Q7.Each attribute represents five symptoms. The illustration is shown below for the Q1

1) The spindle leaves (3rd and 14th)) show drying. At a later stage, stalks become hollow and discoloured.

2) Acervuli (black fruiting bodies) grow on peel and nodes. After dividing open the diseased stalk, an acid smell emanate.

3) The internal tissues are reddened with intermingle transverse white spots.

The sample Disease Names taken for the research is given below

- Grassy_shoot
- Leaf_scald_disease
- Mosaic_disease
- Pokkahboeng
- Red_rot
- Red_striped_disease
- Rust
- Smut
- Sugarcane_yellow_leaf_disease
- Wilt

The Sample of 100 instances is taken as Training data for the sugarcane. The below-mentioned values are taken as Q1, Q2, Q3, Q4, Q5, Q6,Q7, and the disease name are shown in Table 7.

Table 7 Attributes with the Disease Name for Sugarcane

Attributes							Disease
Q1	Q2	Q3	Q4	Q5	Q6	Q7	
5	5	2	1	2	3	3	Red_striped_disease
1	1	2	1	1	5	2	Smut
4	1	2	1	1	2	4	Wilt
4	1	2	2	3	4	4	Smut
3	1	4	2	5	4	5	Pokkahboeng
1	3	5	1	1	2	4	Red_rot
1	3	1	4	5	3	3	Mosaic_disease
3	1	3	3	2	5	1	Smut
4	5	3	2	3	5	5	Sugarcane_yellow_leaf_disease

The Table-8 depicts the Sugarcane Training and Testing Error measures. It is compared with C4.5, SVM , ANN and DA algorithms

Table 8 Testing Error Measures for Paddy

Algorithms	C4.5	SVM	ANN	DA
No of instances	100	100	100	100
Correctly classified	59	91	64	100
Incorrectly classified	41	9	36	0
Kappa Statistic	0.5408	0.899	0.5972	1
Mean absolute error	0.0927	0.1604	0.1132	0
Root mean squared error	0.2153	0.2719	0.2238	0
Relative absolute error	51.827%	89.734%	63.3265%	0.0001%
Relative Squared error	72.0129%	90.976%	74.8593%	0.0002%
Coverage of cases	100%	100%	89%	100%
Mean rel. region size (0.95 level)	24.8%	82.3%	52.6%	10%

The proposed DA algorithm is compared with C4.5, SVM, ANN and the performance metrics are compared with the existing algorithms. The comparison is given in the Table-9

Table 9 Comparative Measures of Each Technique for Sugarcane

Algorithms	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
C4.5	0.59	0.05	0.607	0.59	0.575	0.95
SVM	0.91	0.014	0.92	0.91	0.912	0.99
ANN	0.64	0.037	0.654	0.64	0.621	0.905
DA	1	0	1	1	1	1

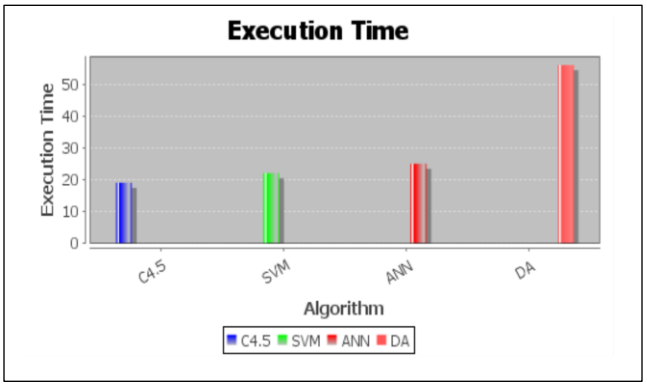

Figure 9 Execution Time Comparison

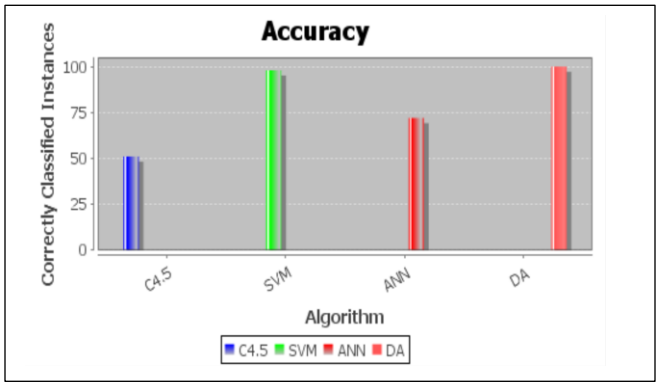

Figure 10 Accuracy Comparison

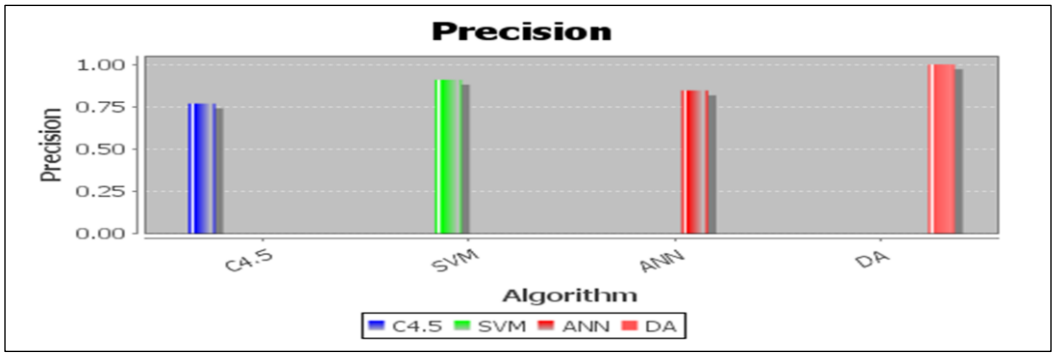

Figure 11 Precision Comparison

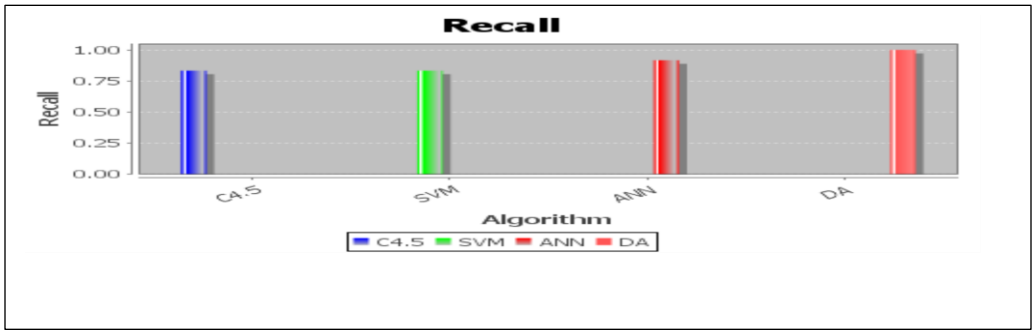

Figure 12 Recall Comparison

The results are visually shown through the figures 9,10,11 and 12. In Figure 9, the execution time of the proposed algorithm is compared with the previous algorithms. As well as in Figure 10, 11 and 12, the accuracy, precision and recall are compared. Through these performance metrics, it is proven that the proposed algorithms are betted then the existing algorithms in terms of performance. Disadvantage of SVM is it's hard to decide best parameters when training information is not linearly divisible.

6 Conclusion

Agriculture is an innovative area of research and is anticipated to develop in the upcoming. There is a lot of work to be done in this future and mind-blowing research area. A multidisciplinary method of combining computation with agriculture would assist in predicting/manage crops efficiently. This paper analyzed the various feature sets of different data mining classifiers to forecast crop diseases effectively.

Use of LDA Models for Crop Disease Predictions Current work on banana, paddy and sugarcane forecasts has emerged as a useful tool for future prognosis modelling. Experimental methods are integrated with dynamic models, which are acceptable methods to comprehend systems' dynamics better. To date, only many regression and neural networks have been comprehensively utilized to predict plant diseases in different parts of the world. The high predictive accuracy of the latest ML techniques such as LDA, as shown in the current study, would enable control systems that make more proficient predictions and reduce yield losses.

References

[1] Anuradha Badage. "Crop Disease Detection using Machine Learning: Indian Agriculture", International Research Journal of Engineering and Technology (IRJET), Vol. 5, no. 9, pp. 866-869, 2018.

[2] P. Revathi, R. Revathi and Dr.M.Hemalatha. "Comparative Study of Knowledge in Crop Diseases Using Machine Learning Techniques", International Journal of Computer Science and Information Technologies, Vol. 2, no. 5, pp. 2180-2182, 2011.

[3] B. Devika, B. Ananthi. "Analysis Of Crop Yield Prediction Using Data Mining Technique To Predict Annual Yield Of Major Crops", International Research Journal of Engineering and Technology, Vol. 5, no. 12, pp. 1460-1465, 2018.

[4] D.S.Mokashi, P.M.Ghodke, A.S.Chadchankar. "Crop Disease Prediction Using Data Mining Method", International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences, Vol. 6, no. 5, pp. 229-230, 2014.

[5] A.Nithya, Dr.V.Sundaram. "Wheat disease identification using Classification Rules", International Journal of Scientific & Engineering Research, Vol. 2, no. 9, 2011.

[6] U. Ayub and S. A. Moqurrab. "Predicting crop diseases using data mining approaches: Classification", 1st International Conference on Power, Energy and Smart Grid (ICPESG), 2018.

[7] Rakesh Kaundal, Amar S Kapoor and Gajendra PS Raghava. "Machine learning techniques in disease forecasting: a case study on rice blast prediction", BMC Bioinformatics, Vol. 7, 2006.

[8] Jagadeesh D.Pujari, Rajesh Yakkundimath and Abdulmunaf. Syedhusain Byadgi. "SVM and ANN Based Classification of Plant Diseases Using Feature Reduction Technique", International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 3, no. 7, pp. 7-14, 2007.

[9] Anuradha, Kuldeep Kaswan, Sugandha Singh. "Two Stage Classification Model for Crop Disease Prediction", International Journal of Computer Science and Mobile Computing, Vol. 4, no. 6, pp. 254-259, 2015.

[10] Cora B. Perez-Ariza. "Prediction of Coffee Rust Disease Using Bayesian Networks", 6th European Workshop on Probabilistic Graphical Models, pp. 259-266, 2012.

[11] Savita N. Ghaiwat, Parul Arora. "Detection and Classification of Plant Leaf Diseases Using Image processing Techniques: A Review", International Journal of Recent Advances in Engineering and Technology, Vol. 2, no. 3, pp. 2347-2812, 2014.

[12] Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong Li. "Image Recognition of Plant Diseases Based on Backpropagation Networks", 5th International Congress on Image and Signal Processing (CISP), 2013.

[13] Paul PA, Munkvold G.P., "Regression and Artificial Neural Network Modeling for the Prediction of Gray Leaf Spot of Maize", Phytopathology, Vol. 95, no. 4, pp. 388-396, 2005.

[14] M. G. Hill, P. G. Connolly, P. Reutemann and D. Fletcher. "The use of data mining to assist crop protection decisions on kiwifruit in New Zealand", Computers and electronics in agriculture, Vol. 108, pp. 250-257, 2014.

[15] D. C. Corrales, J. C. Corrales and A. Figueroa-Casas. "Towards detecting crop diseases and pest by supervised learning", Ingenieríay Universidad, Vol. 19, no. 1, pp. 207-228, 2015.

[16] Uno Y, Prasher SO, Lacroix R, Goel PK, Karimi Y, Viau A, Patel RM. "Artificial neural networks to predict corn yield from compact airborne spectrographic imager data", Comput Electron Agr., Vol. 47, no. 2, pp. 149-161, 2005.
[17] El-Telbany M, Warda M, El-Borahy M. "Mining the classification rules for egyptian rice diseases", Int., Arab J., Inf., Technology, Vol. 3, no. 4, pp. 303-307, 2006.
[18] Diriba Z, Borena B. "Application of data mining techniques for crop productivity prediction", HiLCoE Journal of Computer Science and Technology, Vol. 1, pp. 151-155, 2013.
[19] Ashwinirani, Vidyavathi BM. "Ameliorated methodology for the design of sugarcane yield prediction using decision tree", Compusoft - An International Journal of Advanced Computer Technology, Vol. 4, no. 7, pp. 1882-1889, 2015.
[19] Veenadhari S, Mishra B, Singh CD. "Soyabean productivity modelling using decision tree algorithms", International Journal of Computer Applications, Vol. 27, no. 7, pp. 11-15, 2011.
[20] Prashanth Gupta. "Decision Trees in Machine Learning ", Towards Data Science, 2017. Available online: https://towardsdatascience.com/
[21] Revathy Rathinasamy, Lawrance Raj. "Classifying crop pest data using C4.5 algorithm", IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2018.
[22] N. Gandhi, L. J. Armstrong, O. Petkar and A. K. Tripathy. "Rice crop yield prediction in India using support vector machines", 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2016.
[23] Gniewko Niedbała. "Simple model based on artificial neural network for early prediction and simulation winter rapeseed yield", Journal of Integrative Agriculture, Vol. 18, no. 1, pp. 54-61, 2019.
[24] Khan, Firoz. "Detecting malicious URLs using binary classification through ada boost algorithm", International Journal of Electrical & Computer Engineering, Vol. 10, no. 1, 2020.
[25] Kandhasamy, J. Pradeep. "Diagnosis of diabetic retinopathy using multi level set segmentation algorithm with feature extraction using SVM with selective features", Multimedia Tools and Applications, Vol. 79, pp. 10581-10596, 2019.
[26] Khan, Firoz. "A digital DNA sequencing engine for ransomware detection using machine learning", IEEE Access, Vol. 8, pp. 119710-119719, 2020.
[27] Ramasamy, Lakshmana Kumar, Seifedine Kadry, and Sangsoon Lim. "Selection of optimal hyper-parameter values of support vector machine for sentiment analysis tasks using nature-inspired optimization methods", Bulletin of Electrical Engineering and Informatics, Vol. 10, no. 1, pp. 290-298, 2020.

[28] Sornalakshmi, M. “An efficient apriori algorithm for frequent pattern mining using mapreduce in healthcare data”, Bulletin of Electrical Engineering and Informatics, Vol. 10, no. 1, 2020.

[29]NH Niloy MAI Navid, "Data Mining Algorithm on Fuzzy Weighted Association Rules",International Research Journal of Multidisciplinary Science & Technology (IRJMRS),Vol.1,no.7,pp.480-488,2016.

[30] S. Suganya R. Karpagam,"Applications Of Data Mining And Algorithms In Education – A Survey",International Journal Of Innovations In Scientific And Engineering Research (IJISER),Vol.3,no.4,pp.38-46,2016**.**

.

Biographies

M.C.S. Geetha is currently working as an Assistant Professor in theDepartment of Computer Applications, Kumaraguru College of Technology, Coimbatore andalso pursuing part-time PhD (Computer Science) at Avinashilingam University for Women,Coimbatore. She has 15 years of teaching experience with five years of research work. Shereceived her Master of Computer Applications (MCA) degree at P.S.G.R. Krishnnammal Collegefor Women, Coimbatore, India. She received MPhil (Computer Science) at BharathiyarUniversity. She has published many papers in international journals. Her research interest includedata mining and data analytics.

I. Elizabeth Shanthi is currently working as an Associate Professor in Computer Science at Avinashilingam University for Women, Coimbatore. She has 29 years of teaching experiencewith ten years of research work. She has quite a number of publications at her credit. Her areas ofinterests include data mining, information retrieval, object-oriented data bases, cloud computingand soft computing.