International Journal of Engineering & Technology, 7 (3.34) (2018) 908-912

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Crop Yield Prediction from Soil Parameters through Neupper Rule Established Algorithm

[1]S.Manimekalai, [2]K.Nandhini

[1]PhD Research scholar, PG and Research Department of Computer Science, Chikkanna Government Arts College Tirupur, Tamilnadu, India

[2]Assistant Professor, PG and Research Department of Computer Science, Chikkanna Government Arts College, Tirupur, Tamilnadu, India

**Corresponding Author E-mail: [1]manimekalaipari@gmail.com*

Abstract

Agriculture is considered as the backbone of the Indian economy. Soil classification is used to maintain and enhance its productivity, to avoid soil degradation and to overcome environmental damage. Repeated Incremental Pruning to Produce Error Reduction (Ripper) was used for crop yield prediction. Though the computational cost of this method is very much decreased, the accuracy is not much increased for prediction of crop yield. Artificial Neural Network (ANN) was widely used technique for yield prediction of crop. But with a large number of hidden layers and nodes, the training time will increase and leads to over fitting of networks. Sometimes, it comes out with hardly interpretation results. Counter Propagation ANN (CP-ANN) was a technique which combining features of both supervised and unsupervised learning technique which effectively predicted the crop yield. However, the prediction accuracy of this method is still poor. In order to overcome the above problems and to improve the crop yield prediction accuracy, a Neupper rule based algorithm is proposed in this paper. The proposed algorithm combines ANN and Ripper classifier algorithms. The soil parameters such as phosphorus (P), potassium (K), sulphur (S), calcium (C), magnesium (Mg), zinc (Zn), copper (Cu), iron (Fe) and manganese (Mn) are given as input to ANN which returns weight of each soil parameter. Then, a decision tree is constructed based on weight values. After this, rules are generated based on the building and optimization stage of Ripper algorithm. The proposed Neupper rule based algorithm improves the accuracy of crop yield prediction by combining ANN and Ripper. Finally, the experiments are carried out in terms of accuracy, precision and Root Mean Square Error (RMSE) to prove the effectiveness of the proposed crop yield prediction method.

Keywords: *Agriculture; crop yield prediction; feature selection; classification.*

1. Introduction

Crop yield prediction [1] is more significant as it can support decision makers in agricultural sector. It also assists in identifying the relevance of attributes which significantly affect the crop yield. In earlier days, the yield prediction was carried out only by the experience of the farmers. The improved technology covers many ways that leads to know the productivity of crop. The application of data mining techniques is one of the best ways to provide the solution for this crisis. Due to increasing population, the food demands are also increased. Crop models and decision tools are increasingly used in agricultural fields to improve production competence.

Crop yield predict analysis requires a model of how crops respond to soil factors. Various data mining techniques have been used for crop yield prediction. Ripper [2] is one of the supervised classification approaches where the training data was distributed randomly in two sets are growing set and pruning set. In growing set, each rule keeps on growing until no information gain is possible is further. Then in pruning step, unnecessary terms were eliminated. But the crop yield prediction accuracy of Ripper is low. ANN [3] was an efficient technique for crop yield prediction. There are three layers such as input layer, hidden layer and output layer in ANN. The input neurons got the input from user, then in the hidden neurons the input was converted into (-1,1) interval. Finally the output neurons represented the attribute values that were predicted. When increasing the number of hidden layers, the training time for crop yield prediction gets increased. CP-ANN [4] was similar to Self Organizing Maps (SOM) due to the fact that an output layer was appended to SOMs input layer. It attained non linear class limit separation. However, the prediction accuracy of crop of this method is low.

In order to improve the prediction accuracy of crop yield, a Neupper rule based algorithm is proposed in this paper. In the Neupper rule based algorithm, both ANN and Ripper classifiers are combined together. Initially, input neurons of ANN get various soil parameters such as phosphorus (P), potassium (K), sulphur (S), calcium (C), magnesium (Mg), zinc (Zn), copper (Cu), iron (Fe) and manganese (Mn). The output neurons of ANN return the weights of each soil parameter. Then based on the weight values of soil parameters decision tree is constructed through Ripper classifier procedure. The Ripper procedure has two steps they are build rule step and prune rule step. In the build rule step, the rules are built by adding conditions to the rule until the weight value of soil parameter is greater than user defined threshold value. Then, in the prune rule step the unnecessary rules are pruned. Based on the rules, the yield of rice, wheat and maize is predicted.

2. Literature Survey

Various data mining techniques [5] were analyzed for prediction of crop yield by considering an important factor called soil. Data mining techniques were applied to convert huge data into technologies and make them available to the farmers. The huge amount of data was used to mine volume of knowledge which can be useful for farmers and decision makers to take effective and prompt decision. The data were collected from Agriculture University and then the soil were classified based on soil texture, color, drainage and its alkalinity. The fertility of soil was classified by using K-means, Random Forest and Apriori algorithm.

A system [6] was presented using data mining techniques for the analysis of soil behavior and the prediction of crop yield. This system predicted the categorization of the analyzed datasets. Thus the predicted category indicating the yield of crops. In this system, the problem of predicting the crop yield was formalized as a classification rule, where Naïve Bayes (NB) and K-Nearest Neighbor (KNN) were used. NB handled an arbitrary number of variables whether they are continuous or categorical. The class of the variables is assigned based on the probability of the variable. KNN made predictions based on the outcome of the K neighbors closest to the center point. This system is not applicable for larger dataset of 1GB or more.

A web based intelligent system [7] was presented for prediction of apricots yields using Artificial Neural Network (ANN). The main intention of this system is to determine the possibilities for using ANN to predict the apricots yields. This system gets the input of different parameters such as amount of fertilizers, beginning of the harvest, length of the shoots, fruit mass and thickness of shoots. Moreover, a web based application was created that displays the final results obtained through neural network. The Plan, Do, Check, Act method was used in order to ensure the continual and control improvement of the process. However, this system lacks visualization and an overview of the measures that contain input values. Agricultural crop yield prediction using artificial neural network approach

In this paper [8] the approach used is Artificial Neural Network (ANN) to predict the crop yield by sensing different features of soil and features related to atmosphere. The features considered in this approach are organic carbon, nitrogen, phosphate, sulphur, nitrogen, type of soil, PH, potassium, iron, calcium, manganese, magnesium, temperature, depth, humidity and rainfall. These parameters were given as input to ANN. The ANN was inspired by animal central neuron systems. It was capable of predicting the crop yield.

A decision support system prototype [9] was proposed for rice crop yield prediction. The decision support system was developed as Graphical User Interface (GUI) in Java using NetBeans tool and Microsoft office Access database for the ease of farmers and decision makers. In this system, a user selected the range for the parameters and ran the model in order to predict the yield for the study area. The interface allowed the selection of a range of precipitation, minimum temperature, average temperature, maximum temperature and reference crop evapo transpiration as part of input data set. After processing using the past climate database, the model predicted the expected class of yield viz., low, moderate or high as part of output dataset. The ranges of the parameters were calculated based on the historic data of the study area.

A novel framework called eXtensible Crop Yield Prediction Framework (XCPF) [10] was proposed for agricultural crop yield prediction. This framework was designed to be dynamic and flexible so that it was used for prediction of crop yield of various crops. It has the provision for crop selection, independent variable selection, dependent variable section, dataset selection, and pre-processing and crop yield prediction. Finally, the system was integrated with a management information system that can help in expert decision making towards precision agriculture.

3. Proposed Methodology

In this section, the proposed Neupper rule based algorithm for crop yield prediction has been described in detail. Initially, various soil parameters are collected such as phosphorus (P), potassium (K), sulphur (S), calcium (C), magnesium (Mg), zinc (Zn), copper (Cu), iron (Fe) and manganese (Mn). This is given as input to Neupper rule based algorithm which predicts the yield of rice, wheat and maize.

3.1 Neupper Rule based Algorithm for Crop Yield Prediction

The collected soil parameters are fed into the ANN which consists of three layers. They are input layer, hidden layer and output layer. The soil parameters are considered as $X_1, X_2, \ldots X_n$. The hidden layer H_i is defined as:

$$H_i = \sum_{i=1}^{n} w_i X_i + b \quad (1)$$

Where, w_i represents the weights of input layer and X_i represents the soil parameters and b denotes the bias. The hidden layer of ANN is defined as tan-sigmoid transfer function

$$Y_i = f(H_i) = \frac{2}{1+e^{-2H_i}} - 1 \quad (2)$$

The output layer of ANN is defined by the following equation:

$$o_i = f\left(\sum_{i=1}^{n} w_h Y_i + b\right) \quad (3)$$

Where, o_i denotes the output neurons value, $f(x)$ denotes the transfer function, w_h denotes the weights values of the hidden layer, X_i denotes the input soil parameter and b denotes the bias value.

The ANN model is trained for p number of soil features. For the $i^{t□}$ neuron in output layer, if output $t_i \neq o_i$, then for the $p^{t□}$ observation, the error signal for the $i^{t□}$ neuron in output layer is given as:

$$E = \frac{1}{2}\left(t_i^p - o_i^p\right)^2 \quad (4)$$

Where, t_i denotes the desired target output and o_i denotes the actual observed output from the system, for $i^{t□}$ neuron in the output layer. The weight values are updated as:

$$W_{ij_{new}} = W_{ij_{old}} + \delta(t_i - o_i)\, X_i \quad (5)$$

where, $W_{ij_{new}}$ denotes the synaptic new weight to $i^{t□}$ neuron in the output layer from the $j^{t□}$ neuron in the previous layer, $W_{ij_{new}}$ denotes the synaptic old weight to $i^{t□}$ neuron in the output layer from the $j^{t□}$ neuron in the previous layer and δ denotes the learning parameter. The weights are updated until complete set of training set is utilized for training of the network. A new weight values are need to find so that the error value comes or closer to minimum.

The weight values of input layer w_h and weight values of hidden layer w_i are represented in a matrix form which is denoted as W.

$$W = [w_i w_h] \quad (6)$$

The weight values represent the importance of soil feature for prediction of crop yield. The soil feature which has high weight value forms the root of a decision tree. Ripper is an effective technique found in decision tree algorithms that implements a separate and conquer strategy to separate the positive and negative instances. The main intend of this algorithm is to cover all or maximum number of positive instances with no negative instances or minimum number of negative instances. If at least one rule is fired from a new instance, it is classified as positive otherwise it is classified as negative. Ripper is based in association rule with reduced error pruning.

Initially, the training data is divided into a growing set and a pruning set. An initial rule set is formed that overfits the growing set, using some heuristic method. The lengthy rule set is then repeatedly simplified by applying one of the pruning operators which would be used to delete any single rule or any single condition. At each stage of simplification process, the pruning operator chosen is the one set that yields the greatest reduction of error on the pruning set. The simplification process ends when weight value of output layer of ANN is greater than the user defined threshold value. Based on the rules the yield of rice, wheat and maize are predicted.

3.2. Neupper Rule based Algorithm

Input: Soil features, threshold

Output: Rules

//ANN
Step 1: Assign soil parameters for ANN classifier
Step 2: Apply the equation (1) to hidden layer
Step 3: The output layer of ANN is processed using equation (2 & 3)
Step 4: The error signal for i^{th} neuron is calculated using equation (4)
Step 5: The weight of each neurons are updated using equation (5) It leads the error is closer to minimum.
//Ripper
Step 6: Make a root node with the soil feature which has the highest weight values
Step 7: Construct the tree based on the weight values of soil feature
Step 8: Create an empty rule
Step 9: Add conditions as long as the final weight value of ANN is greater than threshold value.
Step 9a: Build rules with accuracy is equal to 1 (if possible)
Step 10: Prune the rule immediately using reduced error pruning
Step 11: Measure pruning by using

$$W(R) = \frac{(pos - neg)}{(pos + neg)}$$

Where, $W(R)$ denotes the pruning measure, pos denotes the positive instances covered by the rule in validation set and neg denotes the negative instances covered by the rule in validation set.
Step 13: Pruning starts from the last condition add to the rule
Step 13a: Create rules that covers some negative instances (accuracy<1).
Step 14: Apply global optimization strategy
Step 14a: Replace each rule in the decision list with two new ones called revised and replacement rules
Step 14b: Revised rule starts with the antecedent of the original rule and adds a new feature
Step 14c: Replacement rule learns a new rule and starts with an empty antecedent
Step 14d: Choose one of the rule from the three rules which has the shortest description length and eliminate the other two rules.

From the above Neupper rule based algorithm, rules are generated which defines the yield of different crops such as rice, wheat and maize using soil parameters.

4. Result and Discussion

In this section, the existing and proposed methods for crop yield prediction are analyzed by the experimental conclusions. The methods compared by the metrics such as accuracy, precision, recall and f-measure. For the experimental purpose, different soil parameters such as phosphorus (P), potassium (K), sulphur (S), calcium (C), magnesium (Mg), zinc (Zn), copper (Cu), iron (Fe) and manganese (Mn) are considered. The dataset with these parameters and its values are used in the existing and proposed algorithms to find the accuracy of Neupper.

4.1. Accuracy

Accuracy is defined as the proportion of the true outcomes (both true positives and true negatives) among the sum of cases observed. It is calculated as follows:

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

Where, the class label is positive and the yield predicted outcome is positive then it is True Positive
If the class label is negative and the yield predicted outcome is negative then it is True Negative
If the class label is negative and the yield predicted outcome is positive then it is False Positive
If the class label is positive and the yield predicted outcome is negative then it is False Negative

The following Table I shows the comparison of accuracy between Ripper, ANN, CP-ANN and Neupper for three different crops are rice, wheat and maize.

Table I: Neupper Performance Accuracy

Methods	Rice	Wheat	Maize
Ripper	0.5021	0.7211	0.7630
ANN	0.9353	0.9408	0.9133
CP-ANN	0.9352	0.9320	0.9592
Neupper	0.9640	0.9544	0.9748

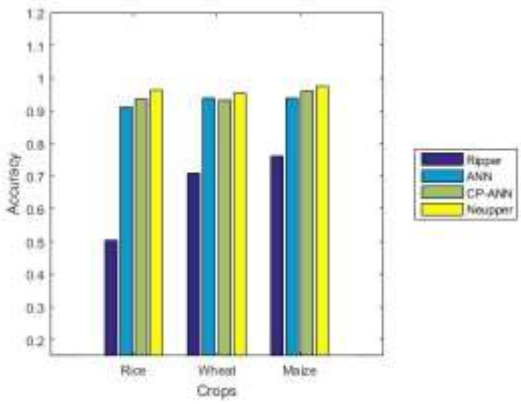

Fig: 1. Accuracy measures

Fig. 1 shows the comparison of accuracy between existing Ripper, ANN, CP-ANN and proposed Neupper based crop yield prediction method. The different crop yield prediction method for three different crops such as rice, wheat and maize is taken in X axis and the accuracy is taken in Y axis. From Fig. 1, it is proved that the proposed Neupper based crop yield prediction method has high accuracy than the other methods.

4.2. Precision

Precision is referred as the computation of exactness or quality, and is measured as follows:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

The following Table II shows the comparison of precision between Ripper, ANN, CP-ANN and Neupper for three different crops is rice, wheat and maize.

Table Ii: Neupper Precision

Methods	Rice	Wheat	Maize
Ripper	0.3783	0.7015	0.7556
ANN	0.9038	0.9335	0.8879
CP-ANN	0.9193	0.9250	0.9671
Neupper	0.9500	0.9400	1

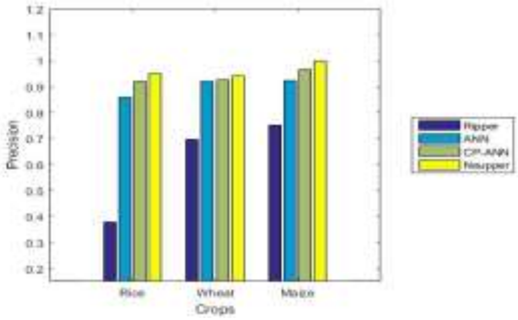

Fig2: Measurement of Precision

Fig. 2 shows the comparison of precision between existing Ripper, ANN, CP-ANN and proposed Neupper based crop yield prediction method. The different crop yield prediction method for three different crops are rice, wheat and maize is taken in X axis and the precision is taken in Y axis. From Fig. 2, it is proved that the proposed Neupper based crop yield prediction method has high precision than the other methods.

4.3. Recall

Recall is defined as the number of true positives divided through the sum of elements which successfully belong to the positive class. It is denoted as follows:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

The following Table III shows the comparison of recall between Ripper, ANN, CP-ANN and Neupper for three different crops is rice, wheat and maize.

Table III: Neupper Recall

Methods	**Rice**	**Wheat**	**Maize**
Ripper	0.8775	0.7620	0.7725
ANN	0.9482	0.9477	0.9253
CP-ANN	0.9895	0.9612	0.9739
Neupper	0.9930	0.9600	0.9900

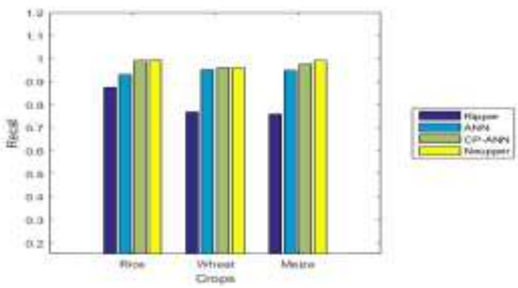

Fig. 3: Measurement of Recall

Fig. 3 shows the comparison of recall between existing Ripper, ANN, CP-ANN and proposed Neupper rule based crop yield prediction method. The different crop yield prediction method for three different crops are rice, wheat and maize is taken in X axis and the precision is taken in Y axis. From Fig. 3, it is proved that the proposed Neupper based crop yield prediction method has high recall than the other methods.

4.4. F-Measure

F-measure computes the mutual value of precision and recall as the harmonic mean of precision and recall. It is computed as follows:

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The following Table IV shows the comparison of f-measure between Ripper, ANN, CP-ANN and Neupper for three different crops is rice, wheat and maize.

Table IV: Comparison of F-measure

Methods	Rice	Wheat	Maize
Ripper	0.5287	0.7305	0.7639
ANN	0.9255	0.9405	0.9062
CP-ANN	0.9531	0.9427	0.9705
Neupper	0.9599	0.9499	0.9950

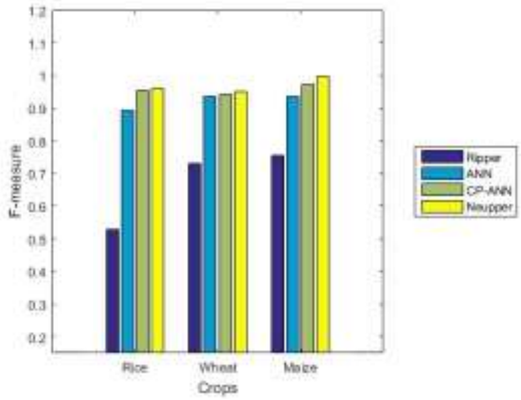

Fig. 4: Measurement of F-measure

Fig. 4 shows the comparison of f-measure between existing Ripper, ANN, CP-ANN and proposed Neupper based crop yield prediction method. The different crop yield prediction method for three different crops are rice, wheat and maize is taken in X axis and the f-measure is taken in Y axis. From Fig. 4, it is proved that the proposed Neupper based crop yield prediction method has high f-measure than the other methods.

5. Conclusion

In this paper, crop yield prediction is improved by combining ANN and Ripper algorithm. Various soil parameters phosphorus, potassium, sulphur, calcium, magnesium, zinc, copper, iron and manganese are considered for prediction of crop yield. The collected soil features are given as input to ANN which returns the weight values. This weight values are used to construct a decision tree. Then rules are generated based on growing phase, pruning phase and optimization phase of Ripper algorithm. In growing phase, conditions are added to the rule until the weight values of ANN is greater than threshold value. In pruning phase, each rule is pruned incrementally. Finally in the optimization phase, delete the rules from the rule set that would increase the description length of whole rule set. Based on the rules, the yield of rice, wheat and maize are predicted based on soil features. The experimental results prove that the proposed Neupper rule based algorithm has high accuracy, precision, recall and f-measures than the other conventional methods.

References

[1] V. Lamba, and V. S. Dhaka, "Wheat yield prediction using artificial neural network and crop prediction techniques (A Survey)", International Journal for Research in Applied Science and Engineering Technology, vol. 2, pp. 330-341, 2014.

[2] S. Vijayarani, and M. Divya, "An efficient algorithm for generating classification rules", International Journal of Computer Science and Technology, vol. 2, no. 4, pp. 512- 515, 2011.

[3] S. Khairunniza-Bejo, S. Mustaffha, and W. I. W. Ismail, "Application of artificial neural network in predicting crop yield: A review", Journal of Food Science and Engineering, vol. 4, no. 1, pp. 1-9, 2014.

[4] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques", Computers and Electronics in Agriculture, vol. 121, pp. 57-65s, 2016.

[5] P. K. Devi, and S. Shenbagavadivu, "Enhanced Crop Yield Prediction and Soil Data Analysis using Data Mining", International Journal of Modern Computer Science (IJMCS), vol. 4, no. 6, pp. 1-7, 2016.

[6] M. Paul, S. K. Vishwakarma, and A. Verma, "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach", IEEE 2015 International Conference on Computational Intelligence and Communication Networks (CICN), pp. 766-771, 2015.

[7] M. Blagojević, M. Blagojević, and V. Ličina, "Web-based intelligent system for predicting apricot yields using artificial neural networks", Scientia Horticulturae, vol. 213, pp. 125-131, 2016.

[8] S. S. Dahikar, and S. V. Rode, "Agricultural crop yield prediction using artificial neural network approach", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, vol. 2, no. 1, pp. 683-686, 2014.

[9] N. Gandhi, L. J. Armstrong, and O. Petkar, "Proposed decision support system (DSS) for Indian rice crop yield prediction", In Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 2016 IEEE, pp. 13-18, 2016.

[10] A. Manjula, and G. Narsimha, "XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction", 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), IEEE, pp. 1-5, 2015.