

A Dynamic Multi-class Neupper Classification for Multiple Crop Yield Prediction

S. Manimekalai, K. Nandhini

Abstract: In agricultural applications, the most essential task is predicting the crop yield to classify the yield productivity over a certain interval of harvesting. The state-of-the-art classifiers are used for predicting the yield quality of any one crop whereas it takes more time to simultaneously train multiple types of crops for predicting their yield quality. For a specific crop yield prediction using soil parameters, a Krill-Herd (KH)-based feature selection with Dynamic Neupper (DNeupper) rule-based classifier has been proposed. However, multiple types of crops were not simultaneously predicted within a single classifier since it creates a multi-class classification problem. Hence in this article, KH with Dynamic Multi-Class Neupper (KHDMCNeupper) rule-based classification algorithm is proposed to predict all three crops such as rice, wheat and maize together with increased prediction accuracy. In this model, the most optimal soil parameters for all crops and their relative weights are computed based on KH and Rough Set (RS) theory. Then, these weight values are combined with soil parameters and given as input to the Artificial Neural Network (ANN) which is used to construct a tree in DNeupper classifier. By constructing a tree, the classification rules for all three crops are generated to predict the yield quality. Thus, the proposed classification technique can support simultaneous prediction of multiple crops with high accuracy. Finally, the experimental results show the efficiency of the KHDMCNeupper classifier compared to the KHDNeupper classifier in terms of accuracy, precision, recall and f-measure.

Index Terms: Crop yield prediction, DNeupper rule-based classifier, KH, Multi-class classification

I. INTRODUCTION

The economical system of each state is mainly related to the agricultural activities or yield productivity. Nonetheless, most of the rural lands still yield less yield productivity due to insufficient deployment of agricultural technologies. Usually, the yield quality depends on various aspects such as healthy soils, water quantity and atmospheric conditions which impact on the plant growth level. In India, the factors that affect the crop yield are inadequate water quantity, yielding range and lack of modern agronomic practices. Therefore, there is a need to support agriculture for balancing the industrial and economic growth of the nation.

Revised Manuscript Received on 30 March 2019.

* Correspondence Author

S. Manimekalai*, Department of Data Mining, Chikkanna Government Arts College, Tirupur, India

Dr. K. Nandhini Department of Computer Science in Chikkanna Government Arts College, Tirupur, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

As a result, different data mining techniques have been proposed to support cultivators in order to predict the crop growth level under those various aspects. Many data mining techniques [1] have been proposed on the basis of historical information about weather, water quantity and crop growth level to increase the yield productivity. Most of the crop yield prediction systems are useful for cultivators to make decisions about the soil type, irrigation level and type of crop to be planted. As well, it supports to predict the desired data from a raw dataset for analyzing the yield productivity issues and making an appropriate decision. Over the past few decades, several classification algorithms [2] have been proposed to predict the crop yield quality. Among those, crop yield prediction has been achieved by classifying the soil parameters based on the Neupper rule-based algorithm which is a combination of Artificial Neural Network (ANN) and Ripper classifiers [3]. In addition, a WK-SMOTE algorithm with feature selection algorithms such as PSO and KH [4] were added to solve the class imbalanced data and choose the most optimal soil features when number of soil parameters were increased. Moreover, DNeupper rule-based algorithm [5] was proposed to enhance the crop yield prediction by finding the feature spaces with a large kernel space between clusters and close proximity of the classes. However, all three crops were not simultaneously predicted within a single classifier since it creates a multi-class classification problem. Hence in this paper, a KHDMCNeupper rule-based classification algorithm is proposed to predict all three crops together with increased prediction accuracy. In this model, the relative weights of the most optimal soil parameters using KH algorithm for all crops are computed based on RS theory. These weight values with soil parameters are given as input to the ANN for constructing a decision tree of DNeupper classifier. Based on the constructed decision tree, the classification rules for all three crops are generated and the crop yield quality is also predicted efficiently for all crops simultaneously. The rest of the paper is organized as follows: Section II presents the previous researches related to the crop yield prediction. Section III explains the proposed crop yield prediction. Section IV illustrates the performance effectiveness of the proposed algorithms and Section V concludes the research work.

II. LITERATURE SURVEY

Mustafa et al. [6] analyzed land suitability for different crops based on the multi-criteria decision making approach using remote sensing and GIS. In this analysis, land suitability was analyzed for generating yielding pattern for kharif and rabi seasons in Khergarah tehsil of Agra.

Agricultural and non-agricultural lands were outlined by using Decision Tree Classifier (DTC) and non-agricultural lands were masked for removal from the future analysis. As well, different soil chemical and physical parameters for different crops were evaluated. After that, those were integrated by using the multi-criteria decision making and GIS for generating the land suitability maps for different crops individually. However, it requires the selection of more number of factors such as weather, irrigation facilities, etc., to improve further efficiency of analysis.

Gholap et al. [7] analyzed soil data using classification techniques. In this analysis, an automated system was developed for soil classification based on fertility. Once the fertility class labels were obtained using the developed automated system, a comparative analysis of different classification techniques such as Naive Bayes, J48 and JRip was carried out with aid of data mining tool known as WEKA. Also, the untested soil attributes were predicted using regression algorithms. However, the accuracy of those classifiers was less.

Ramesh & Vardhan [8] analyzed crop yield prediction using different data mining techniques with the aim of generating a user friendly interface that analyses the rice productivity based on the available data for cultivators. In this analysis, the crop yield prediction for the selected region such as east Godavari district of Andhra Pradesh in India was analyzed based on the Multiple Linear Regression (MLR) and density-based clustering technique. Conversely, some noisy data was also considered as useful data that leads prone to overfitting of the data.

Chaudhary et al. [9] proposed an improved Random Forest Classifier (RFC) for multi-class classification. In this method, an attribute evaluator method and an instance filter method were combined with the RFC to solve the multi-class classification problem and improve the performance of RF algorithm. However, the overfitting problem can occur and need to select the number of trees.

Patel & Patel [10] proposed crop prediction framework by using RS theory. In this framework, RS was used to generate the classification rules from 640 sets of agriculture data for crop monitoring. Initially, the collected data were pre-processed and information table was generated. Then, decision table was also generated. After that, the reduction method was applied to find the reduct of the dataset which holds the minimal subset of attributes associated with the class label. Finally, the rules were generated from the reduct by applying LEM2 algorithm. However, this approach was not suitable for dataset with more number of attributes.

Sinivasan & Shanthi [11] proposed a seed yield estimation modeling by using Classification And Regression Trees (CART) in the bio-fuel supply chain. The main of this model was yielding Jatropha plant in the barren lands by identifying the attributes which affects the crop yield and growth characteristics of Jatropha seed may help the cultivators in the decision-making process of agriculture. In this model, CART was used for predicting the yield estimation of Jatropha seed based on the agricultural dataset. However, the classification accuracy was not increased.

Dolas & Joshi [12] proposed a novel approach for soil classification and crop prediction. In this approach, a modified decision tree classifier was trained with C4.5 and

CART by using the optimal soil parameters such as pH, electrical conductivity and exchangeable sodium percentage. Based on this, a type of soil was identified and crops depending on that soil type were also predicted. On the other hand, it requires an additional improvement on crop yield prediction to schedule the irrigation process.

III. PROPOSED METHODOLOGY

In this section, the proposed KHDMCNeupper classification algorithm is explained in brief. At first, available soil parameters such as phosphorus (P), potassium (K), sulphur (S), calcium (Ca), magnesium (Mg), zinc (Zn), copper (Cu), iron (Fe) and manganese (Mn) are collected. After that, WK-SMOTE algorithm with KH algorithm is applied to solve the data imbalance problem and select the most optimal soil parameters to be given to the DNeupper classifier for predicting the all crop yields simultaneously. This proposed model has three steps such as assignment of weights to the soil parameters, construction of decision tree and classification of data. The details of multi-class DNeupper classifier are described in below sub-section.

A. Computation of Weights of Soil Parameters for All Crops

The weights of selected most optimal soil parameters for each crop is computed by using RS theory with a decision matrix $P(p_{ij}, i = 1, 2, \dots, m; j = 1, 2, \dots, n)$ where m denotes the number of crops to be considered and n denotes the number of soil parameters. Each row of the decision matrix is allocated to any one crop and each column to one parameter. Therefore, an element p_{ij} of the decision matrix indicates the quality of i^{th} crop with respect to j^{th} parameter. The relative importance (priority weights) of each parameter is obtained by using dominance-based RS theory as follows:

$$\gamma_{pj} = 1 - \frac{Card(\{a_i \in A | D_{P-p_j}^+(a_i) \subseteq D_P^+(a_i)\})}{Card(A)} \quad (1)$$

In (1), A refers to the normalized matrix of P , $D_P^+(a_i)$ refers to the set of elements dominating a_i with regard to parameter set P and $D_{P-p_j}^+(a_i)$ denotes the set of elements dominating a_i with regard to parameter set after removing the parameter p_j , namely $P - p_j$. Once the value of γ_{pj} is calculated, the weight of each parameter is computed by normalization by using the following formula:

$$w_j = \frac{\gamma_{pj}}{\sum_{j=1}^n \gamma_{pj}} \quad (2)$$

B. Classification using DNeupper Algorithm

The computed weight of each parameter is multiplied with their corresponding parameter values. Therefore, the input of ANN in DNeupper classifier is updated as follows:

$$X_{i_new} = X_i w_j \quad (3)$$

Then, the hidden layer of ANN is defined as:

$$H_i = \sum_{i=1}^n w_i X_{i_new} + b \quad (4)$$

In (4), w_i represents the weights of input layer and X_{i_new} represents the newly updated soil parameters values and b denotes the bias.

The hidden layer of ANN is defined as the following tan-sigmoid transfer function:

$$Y_i = f(H_i) = \frac{2}{1+e^{-2H_i}} - 1 \quad (5)$$

The output layer of ANN is defined as:

$$o_i = f(\sum_{i=1}^n w_h Y_i + b) \quad (6)$$

In (6), o_i denotes the output neurons value, $f(x)$ denotes the transfer function, w_h denotes the weights values of the hidden layer. As well, the weight value of ANN is updated to minimize the error of i^{th} neuron in the output layer as follows:

$$W_{ij_{new}} = W_{ij_{old}} + \delta(t_i - o_i) X_i \quad (7)$$

In (7), t_i denotes the desired target output, $W_{ij_{new}}$ denotes the synaptic new weight to i^{th} neuron in the output layer from the j^{th} neuron in the previous layer, $W_{ij_{old}}$ denotes the synaptic old weight to i^{th} neuron in the output layer from the j^{th} neuron in the previous layer and δ denotes the learning parameter. Based on the output neuron values, the soil parameter with the highest weight value is chosen which acts as root node in the Ripper classifier to construct the decision tree. Once tree is constructed, the classification rules for all crops are generated to predict the yield quality of all crops together using single classifier known as DNeupper classifier.

Algorithm for DNeupper Classifier-based Crop Yield Prediction

Input: Soil parameters such as P, K, S, Ca, Mg, Zn, Cu, Fe and Mn for rice, wheat and maize

Output: Yield quality of all crops

Collect the soil parameters of all crops;
for (each soil parameter of all crops)

{
//Solve class imbalanced data problem using WK-SMOTE algorithm;

Initialize training dataset, $S^{seed}, S^{neighbor} = \{ \}$;

for ($l = 1$ to P)

{
Randomly sample x_p from S^{min} ;

Compute k -nearest minority neighbors of x_p ;

Randomly sample a neighbor as x_q ;

Include x_p, x_q to $S^{seed}, S^{neighbor}$ respectively;

}
Determine augmented kernel matrix;

Solve ANN decision function;

Return predict class of x ;

End

//Regulate the feature space dynamically;

$$Y^{total} = [Y_{opt}^1, Y_{opt}^2, \dots, Y_{opt}^N];$$

$$\Phi(L^{total}) = \frac{1}{2\pi^{(O/2)}\sigma^O} \exp\left(-\frac{\|L^{total} - Y_g^{total}\|^2}{2\sigma^2}\right);$$

Y^{total} & L^{total} are O -dimensional feature row vector representations, σ refers the spread of the Gaussian function and $\Phi(L^{total})$ is a conditional probability of a new feature vector belonging to the class of Y^{total} ;

Select the most optimal parameters using KH algorithms;

//Compute relative weights of each selected

parameters;

$$\gamma_{pj} = 1 - \frac{Card(\{a_i \in A | D_{p-p_j}^+(a_i) \subseteq D_p^+(a_i)\})}{Card(A)};$$

$$w_j = \frac{\gamma_{pj}}{\sum_{j=1}^n \gamma_{pj}};$$

//Update the soil parameter value;

$$X_{i_{new}} = X_i w_j;$$

Assign the new soil parameter values as input to the ANN classifier;

Process the ANN hidden and output layers;

//Update the weight of each neuron to minimize the error;

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

//Obtain the soil parameter with the highest weight as output of ANN;

$$o_i = f(\sum_{i=1}^n w_h Y_i + b);$$

Make a soil parameter with the highest weight as a root node;

Construct a tree based on Ripper algorithm;

Generate the classification rules for all crops;

Predict the yield quality;

}

IV. RESULTS AND DISCUSSION

In this section, the performance of proposed KHDNeupper classifier-based crop yield prediction model is analyzed and compared with existing KHDNeupper based prediction model by using MATLAB 2018a. The dataset used for this experiment is gathered from Tamil Nadu Agricultural University (TNAU), Coimbatore. The collected dataset consists of soil parameters such as P, K, S, Ca, Mg, Zn, Cu, Fe and Mn for wheat, rice and maize. The comparison analysis is made in terms of different metrics like precision, recall, f-measure and accuracy.

A. Precision

It is computed based on the prediction at True Positive (TP) and False Positive (FP) rates.

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

The following Table 1 shows the comparison of precision between proposed and existing crop yield prediction models for three different crops such as rice, wheat and maize.

Table 1: Comparison of Precision

Models	Rice	Wheat	Maize
KHDNeupper	0.980	0.975	0.983
KHDNeupper	0.984	0.980	0.986

A Dynamic Multi-class Neupper Classification for Multiple Crop Yield Prediction

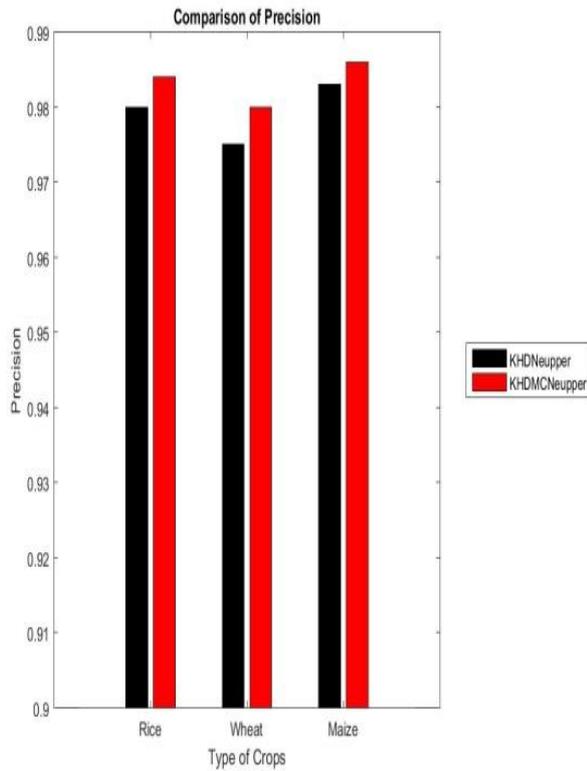


Fig.1: Comparison of Precision

Fig.1 shows the comparison of precision for KHDNeupper and proposed KHDMCNeupper. In the graph, different crops such as rice, wheat and maize are considered in the x-axis and the value of precision is taken in the y-axis. From the analysis, it is observed that KHDMCNeupper classifier has high precision than the KHDNeupper. For example, the precision of KHDMCNeupper for rice is 0.41% higher than KHDNeupper classifier.

B. Recall

It is computed based on the prediction value at TP and False Negative (FN) rates.

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

The following Table 2 shows the comparison of recall between proposed and existing crop yield prediction models for three different crops such as rice, wheat and maize.

Table 2: Comparison of Recall

Models	Rice	Wheat	Maize
KHDNeupper	0.881	0.922	0.965
KHDMCNeupper	0.985	0.983	0.985

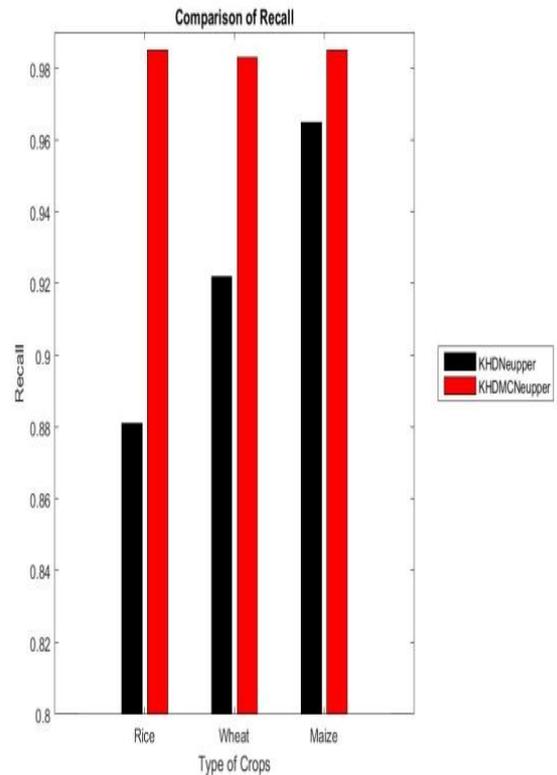


Fig.2: Comparison of Recall

Fig.2 shows the comparison of recall for KHDNeupper and proposed KHDMCNeupper. In the graph, different crops such as rice, wheat and maize are considered in the x-axis and the value of recall is taken in the y-axis. From the analysis, it is observed that KHDMCNeupper classifier has high recall than the KHDNeupper. For example, the recall of KHDMCNeupper for wheat is 6.62% higher than KHDNeupper classifier.

C. F-Measure

It is computed by using both precision and recall as:

$$F - measure = 2 \cdot \left(\frac{Precision \cdot Recall}{Precision + Recall} \right) \quad (10)$$

The following Table 3 shows the comparison of f-measure between proposed and existing crop yield prediction models for three different crops such as rice, wheat and maize.

Table 3: Comparison of F-Measure

Models	Rice	Wheat	Maize
KHDNeupper	0.9855	0.9843	0.9881
KHDMCNeupper	0.9880	0.9870	0.9910

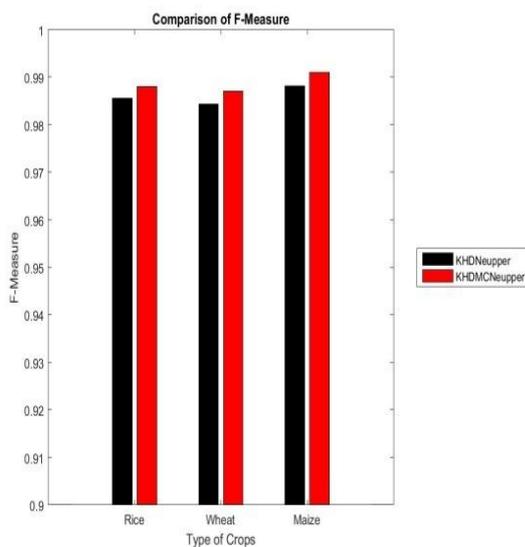


Fig.3: Comparison of F-measure

Fig.3 shows the comparison of f-measure for KHDNeupper and proposed KHDMCNeupper. In the graph, different crops such as rice, wheat and maize are considered in the x-axis and the value of f-measure is taken in the y-axis. From the analysis, it is observed that KHDMCNeupper classifier has high f-measure than the KHDNeupper classifier. For example, the f-measure of KHDMCNeupper for maize is 0.29% higher than KHDNeupper classifier.

D. Accuracy

It is the ratio of TP and True Negative (TN) to the sum of amount of cases examined. It is computed as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

The following Table 4 shows the comparison of accuracy between proposed and existing crop yield prediction models for three different crops such as rice, wheat and maize.

Table 4: Comparison of Accuracy

Models	Rice	Wheat	Maize
KHDNeupper	0.978	0.969	0.982
KHDMCNeupper	0.993	0.990	0.992

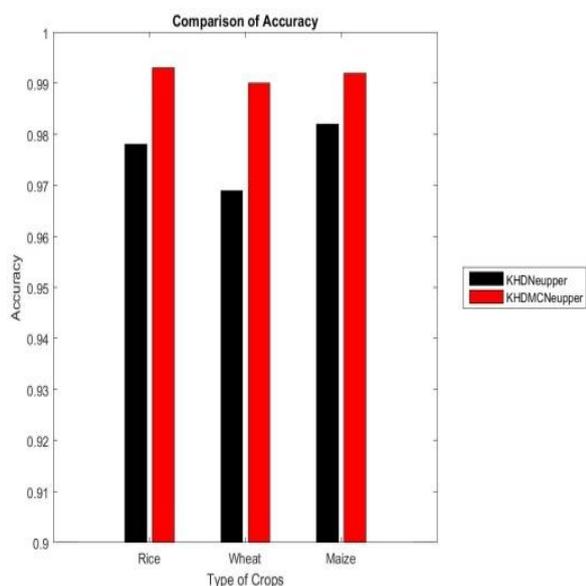


Fig.4: Comparison of Accuracy

Fig.4 shows the comparison of accuracy for KHDNeupper and proposed KHDMCNeupper. In the graph, different crops such as rice, wheat and maize are considered in the x-axis and the value of accuracy is taken in the y-axis. From the analysis, it is observed that KHDMCNeupper classifier has high accuracy than the KHDNeupper classifier. For example, the accuracy of KHDMCNeupper for rice is 1.53% higher than KHDNeupper classifier.

V. CONCLUSION

In this paper, a KHDMCNeupper rule-based classifier is proposed for crop yield prediction. Based on this newly proposed KHDMCNeupper classifier, multiple types of crops can be trained simultaneously using the single classifier to predict their yield quality. Due to this, the training efficiency is increased compared to the training of each type of crop. Also, the prediction accuracy is approximately equal to the classification of each crop individually. As a result, this classifier can be very helpful in real-time applications to predict the yield quality of multiple types of crops at the same time.

REFERENCES

1. P. Priya, U. Muthaiah and M. Balamurugan, "Predicting yield of the crop using machine learning algorithm," *Int. J. Eng. Sci. Res. Technol.*, vol. 7, no. 4, pp. 1-7, 2018.
2. R. Sujatha and P. Isakki, "A study on crop yield forecasting using classification techniques," in *IEEE Int. Conf. Computing Technol. Intell. Data Eng.*, pp. 1-4, 2016.
3. S. Manimekalai and K. Nandhini, "Crop yield prediction from soil parameters through Neupper rule established algorithm," *Int. J. Eng. Technol.*, vol. 7, no. 3.34, pp. 908-912, 2018.
4. S. Manimekalai and K. Nandhini, "An imbalanced soil data classification with optimized features for crop yield prediction", *International conference Artificial Intelligence, Smart Grid and Smart City Applications* at PSG College of Technology on January 2019.
5. S. Manimekalai and K. Nandhini, "A dynamic neupper classification-based crop yield prediction using class imbalanced optimal soil parameters", *Journal of Advanced Research in Dynamical and Control Systems*.
6. A. A. Mustafa, M. Singh, R. N. Sahoo, N. Ahmed, M. Khanna, A. Sarangi and A. K. Mishra, "Land suitability analysis for different crops: a multi criteria decision making approach using remote sensing and GIS," *Res.*, vol. 3, no. 12, pp. 61-84, 2011.
7. J. Gholap, A. Ingole, J. Gohil, S. Gargade and V. Attar, "Soil data analysis using classification techniques and soil attribute prediction," *arXiv preprint arXiv:1206.1557*, 2012.
8. D. Ramesh and B. V. Vardhan, "Analysis of crop yield prediction using data mining techniques," *Int. J. Res. Eng. Technol.*, vol. 4, no. 1, pp. 47-473, 2015.
9. A. Chaudhary, S. Kolhe and R. Kamal, "An improved random forest classifier for multi-class classification," *Inf. Process. Agric.*, vol. 3, no. 4, pp. 215-222, 2016.
10. H. Patel and D. Patel, "Crop prediction framework using rough set theory," *Int. J. Eng. Technol.*, vol. 9, no. 3, pp. 2505-2513, 2017.
11. S. P. Srinivasan and D. S. Shanthi, "A seed yield estimation modelling using classification and regression trees (CART) in the biofuel supply chain," *J. Biomed. Imaging Bioeng.*, vol. 1, no. 1, pp. 8-12, 2017.
12. V. M. Dolas and U. Joshi, "A novel approach for classification of soil and crop prediction," *Int. J. Comput. Sci. Mob. Comput.*, vol. 7, no. 3, pp. 20-24, 2018.

A Dynamic Multi-class Neupper Classification for Multiple Crop Yield Prediction

AUTHORS PROFILE



S. Manimekalai received her BCA in 2007, from Bharathiar University, Coimbatore and MCA in 2011, from Anna University, Coimbatore. She obtained her M.Phil in the area of Data Mining from Bharathiar University, Coimbatore in 2014. At present, she is pursuing her Ph.D in the area of Data Mining in Chikkanna Government Arts College, Tirupur.



Dr. K. Nandhini received her B.Sc., from Bharathiar University, Coimbatore in 1996 and received M.C.A from Bharathidasan University, Trichy in 2001. She obtained her M.Phil in the area of Data Mining from Bharathidasan University, Trichy in 2004. She obtained her Ph.D degree in the area of data mining from Bharathiar University, Coimbatore in 2012. At present she is working as an Assistant Professor at PG & Research Department of Computer Science in Chikkanna Government Arts College, Tirupur. She has published more than 25 research papers in various National and International journals in the area of Data Mining. Her research interest lies in the area of Data Mining and Artificial Intelligence. She is a life member of the professional body Computer Society of India (CSI).