The Novel Technique Based Gaussian Firefly Selection with Modified Fuzzy Cognitive Maps Prediction on Vegetable Plants.

[1]Dr. Sabareeswaran D, [2]Dr. R. Gunasundari, [3]Dr. K.Ramesh, [4]Dr. M. Ramaraj

[124]Assistant Professor,[3]Professor,[123]Department of Computer Applications, [4]Department of CS, CT& IT,
[1234]Karpagam Academy of Higher Education, Coimbatore.

sabaredhandapani@gmail.com

Abstract

The broadest economic zone in developing countries is agriculture and it has the major responsibility for improving the economic growth of the country. Over the past decades, the growth of agriculture is monitoring by using the different data mining techniques. However, the plant yield prediction requires further improvement. Hence in this article, the plant yield prediction is improved by considering the different features such as soil and weather characteristics. Such features are extracted and the most optimal characteristics are chosen by using the Non-Dominated Sorting Firefly (NDSF) and Gaussian Firefly (GF) algorithm. These algorithms solve the Pareto-front issue and movement speed of firefly towards the best global solution. Then, the selected features are classified by using the Modified Fuzzy Cognitive Map (MFCM) algorithm for predicting the growth of plant yield. Finally, the predicted outcomes are broadcasted to the farmers for identifying the causes for plant yield degradation. The experimental results illustrate that the proposed GFFS-MFCM based plant yield prediction achieves high accuracy compared with the other techniques.

Keywords— Plant yield prediction, Soil characteristics, Weather parameters, Non-dominated sorted firefly algorithm, Gaussian firefly algorithm, Modified fuzzy cognitive map, Pareto-front problem.

1. Introduction

Data mining techniques play a vital role in decision making process associated with the agriculture [1]. Such data mining techniques have been developed rapidly in plant yield prediction since agriculture has consists of huge amount of datasets such as plant data, soil data, weather data, etc. Mostly, the yielding depends on the different factors such as climatic changes, geographical conditions, and economic factors [2]. Data mining algorithms such as Transductive Support Vector Machine (TSVM) and Artificial Neural Network (ANN) have predicted plant disease and its causes in previous studies. Furthermore, the plant yield prediction has been achieved by using the Firefly feature selection based Modified Fuzzy Cognitive Maps (FFFS-MFCM) algorithm. These algorithms help cultivators for predicting the plant yield causes for its degradation effectively. However, Pareto-front issue is not solved in the firefly algorithm and the movement speed of fireflies is not effective. Hence in this article, the Firefly algorithm for Feature Selection (FFFS) is improved by Non-Dominated Sorting Firefly (NDSF) algorithm for solving the Pareto-front issue and Gaussian Firefly (GF) algorithm for improving the speed of movement of fireflies towards the best optimal solutions. Once, the features are selected, MFCM approach is introduced for predicting the plant yield and the obtained outcomes are transmitted to the farmers in order to identify the causes for plant yield degradation.

2. Literature survey

Cheng, H., et al. proposed [3] using image analysis and tree canopy characteristics with Artificial Neural Networks in the early yield prediction model. The major objective of this approach was describing the extraction processes of canopy features and learning the relationship between the features and actual yield per tree by utilizing the Back Propagation Neural Network (BPNN). However, the selection of learning rate was more complex.

Papageorgiou, E. I., et al. proposed [4] Model of yield prediction in apples based on Fuzzy Cognitive Maps' (FCM) dynamic impact graph. A data-driven non-linear FCM learning method for categorizing apple yields was suggested in this approach, in which some decision-making algorithms were identified. However, the classification accuracy was less.

Bornn, L., &Zidek, J. V. [5] investigated about how spatial dependence was incorporated into the statistical models for predicting the crop yieldIn this method, by selecting the biophysically dependent explanatory variables and using the spatially defined prior probability distributions, a Bayesian model was developed for crop yield, but the difficulty of the computation was high.

3. Proposed methodology

The suggested enhanced plant yield prediction system is briefly explained in this section.

3.1 Feature Selection Process

At first, various plants, soil and weather images are captured by means of digital camera with the required resolution and the captured images are transmitted through the wireless networks to the image processing units for further processing. The collected images are pre-processed for removing the unwanted noises and the enhanced images are used for segmentation process using Region of Interest (ROI) method which split the images into smaller regions. These smaller regions are further utilized for extracting the features from the plants, soil and weather images. From plant images, texture, geometrical and shape features are extracted. The soil characteristics such as color, texture, moisture, pH value, organic matter, soil depth, etc., are extracted from the soil images. By using the weather images, wind, temperature, humidity and rainfall are extracted.

3.1.1 Non-Dominated Sorted Firefly (NDSF) Algorithm

Initially, N number of fireflies $\{X = (X_1, X_2, \dots, X_N)\}$ positioned randomly in which x_i are distributed among the search or decision space as uniformly and each firefly has objective functions$f_N(X) = \{f_1(X), f_2(X), \dots, f_N(X)\}$. For an optimization problem, the fitness of a solution is proportional to the value of the objective functions. In the Non-Dominated Sorted Firefly (NDSF) algorithm, the population is modified with good solutions for each iteration and the global search functionality is included in the update method for the non-dominated sorting and population crowding distance selection. [6].

Algorithm: NDSF based Feature Selection

1. Initialize the number of fireflies (Soil and Weather features) and objective function $f(X)$
2. Compute objective function using the light intensity of each fireflies $I(x) = \frac{I_s}{d^2}$
3. Define light absorption coefficient γ
4. Compute the attractiveness function $\beta(x) = \beta_0 e^{-\gamma d^2}$
5. $While(t < MaxGeneration)$
6. $For i = 1{:}n$ do
7. $For j = 1{:}n(j \neq i)$ do
8. $If(fireflyjdominatesfireflyi)$
9. Move firefly i towards firefly j
10. Generate a new one if not all constraints are satisfied
11. End If
12. $If(noonedominatesfireflyi)$
13. Find firefly j that is nearest to firefly i
14. $For k = 1{:}n(k \neq j \text{ and } k \neq i)$
15. Find firefly k that is nearest to but on opposite side of i
16. End $For k$
17. Compute the crowding distance D_i by using $D_i = d_{ij} + d_{ik}$
18. Move randomly to a new solution
19. End $For\ k$
20. End $For\ j$

21. End $For i$
22. Update the non-dominated solutions
23. Sort fireflies and find the current best firefly (features)
24. End $While$
25. Obtain the most optimal features

3.1.2 Gaussian Firefly (GF) Algorithm

The pareto-front issue is removed by using NDSF algorithm. However, the speed of movement of fireflies such as convergence speed is not efficiently optimized. Hence, Gaussian Firefly algorithm based feature selection (GFFS) is introduced for improving the convergence speed effectively [7].

Algorithm: GF based Feature Selection

1. Initialize the number of fireflies (Soil & Weather features) and objective function $f(X)$
2. Define light absorption coefficient γ
3. $While(t < MaxGeneration)$
4. $For i = 1{:}n$ do
5. $For j = 1{:}n$ do
6. Compute objective function using the light intensity of each fireflies $I(X) = \frac{I_s}{d^2}$
7. $If(I_i > I_j)$
8. Move firefly i towards j in all d dimensions
9. $Else$
10. Move firefly i towards the best solution in that iteration
11. End if
12. Compute the attractiveness function $\beta(X) = \beta_0 e^{-\gamma d^2}$
13. End $For\ j$
14. End $For i$
15. Sort fireflies and find the current best firefly
16. Define normal Gaussian distribution
17. $For k = 1{:}n$ (all n fireflies)
18. Obtain a random number from defined distribution
19. Apply equation (11) for introducing social behavior
20. Evaluate $newsolution(new\_cost(k))$
21. $If\left(\left(new\_cost(k) < cost(i)\right) \&\& \left(new\_cost(k) < last\_cost\_iteration(k)\right)\right)$
22. Move position of firefly i towards the current best
23. End If
24. End $For\ k$
25. End $While$
26. Obtain the most optimal features

3.2 Feature Classification Process

The most optimal characteristics acquired are then categorized using the Modified Fuzzy Cognitive Map (MFCM) to predict plant yield. [8].

Thus, the weight matrix is stored and FCM is designed to help efficiently predict the yield of the plant. The MFCM model for the proposed approach is shown in Figure 2 and the adjacency connection matrix is given in Table 1.

Table.1 Adjacency Connection Matrix

	Soil Features	Weather Features
Soil Features	Low Yield (0)	High Yield (1)
Weather Features	High Yield (1)	Low Yield (0)

To express the output categories that can be divided into two groups, such as low yield (OP1) and high yield, the Decision Output Definition (DOC) is assigned to (OP2).Such rules are given below:

- If both selected features are in soil category, Then the output becomes low yield (0);
- If the selected features are in both soil and weather category, Then the output becomes high yield (1);
- If both selected features are in weather category, Then the output becomes low yield (0);

Algorithm: MFCM based Classification

1. Initialize the FCM with number of nodes N and weight matrix W
2. Assign the weight parameters τ and δ
3. Compute each node of

$$A_m^{(k)} = f\left(A_m^{(k-1)} + \sum_{n \neq m} A_n^{(k-1)} \cdot W_{nm}\right)$$

4. For number of iterations do
5. Update the weights based on the following equation

$$W_{nm}(k) = \frac{(1-\tau) \cdot W_{nm}^{(k-1)} + \delta \cdot A_n^{(k-1)} \cdot A_m^{(k-1)}}{\left[\sum_{\substack{n=1 \\ n \neq m}} \left((1-\tau) \cdot W_{nm}^{(k-1)} + \delta \cdot A_n^{(k-1)} \cdot A_m^{(k-1)}\right)^2\right]^{\frac{1}{2}}}$$

6. Evaluate the termination criterion

$$|Output^k - Output^{k-1}| < e$$

7. Return the final weight values
8. Predict the plant yield as high or low
9. End for
10. Broadcast the obtained outcomes to the farmers.
11. End

4. Experimental Results

In this section, the performances of Non-Dominated Sorted FireFly for Feature Selection with Modified Fuzzy Cognitive Map (NDSFFFS-MFCM) based plant yield prediction is compared with FireFly for Feature Selection with MFCM (FFFS-MFCM) and Gaussian FireFly for Feature Selection with MFCM (GFFFS-MFCM) in terms of precision, recall, f-measure and accuracy.

4.1 Precision

Precision is measured at true positive and false positive estimation depending on the function classification.

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)}$$

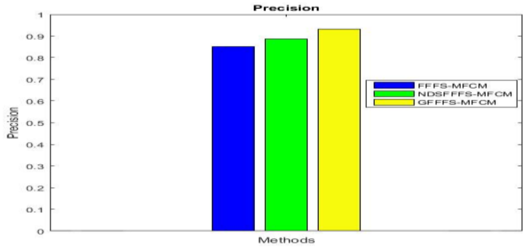
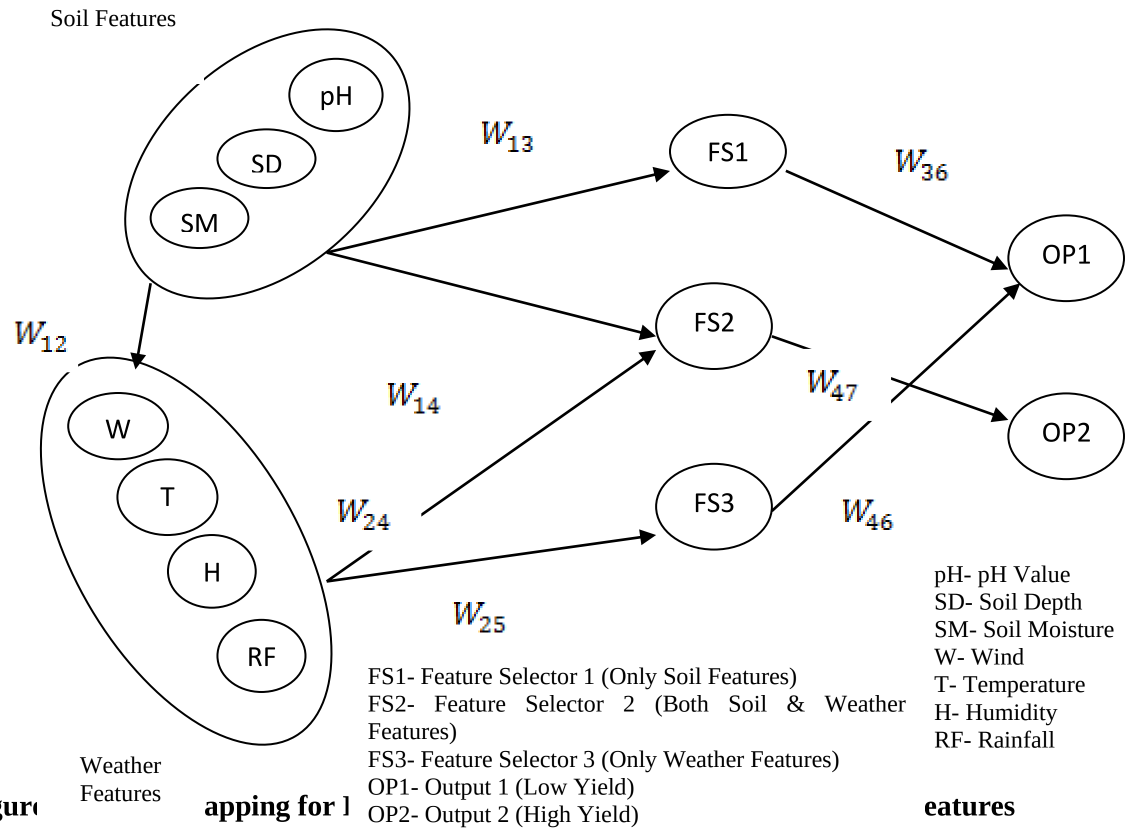


Figure.3 Comparison of Precision

Figure 3 shows that the comparison of precision. From the graph, it is observed that the precision of GFFFS-MFCM increases compared with the other plant yield prediction approaches.



Figur apping for l eatures

4.2 Recall

Based on the function classification, recall is determined on true positive and false negative predictions.

$$Precision = \frac{TruePositive\ (TP)}{TruePositive(TP) + FalseNegative\ (FN)}$$

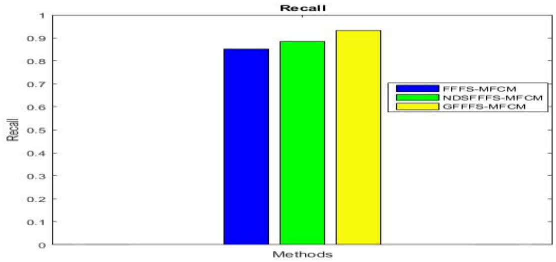


Figure.4 Comparison of Recall

Figure 4 shows that the comparison of recall. From the graph, it is observed that the recall of GFFFS-MFCM increases compared with the other plant yield prediction approaches.

4.3 F-Measure

F-measure is computed by using the values of both precision and recall as follows:

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

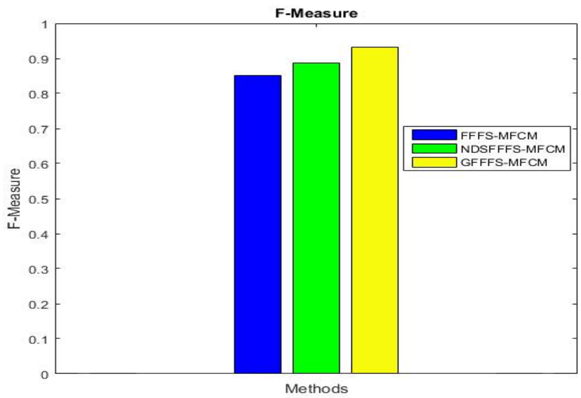
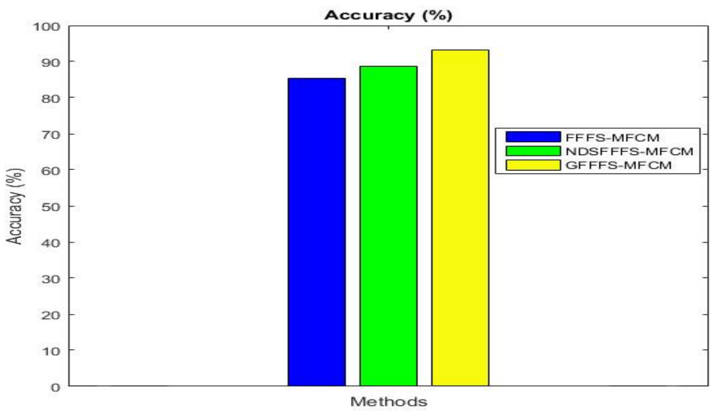


Figure.5 Comparison of F-Measure

Figure 5 shows that the comparison of f-measure. From the graph, it is observed that the f-measure of GFFFS-MFCM increases compared with the other plant yield prediction approaches

4.4 Accuracy

The accuracy of the total number of cases examined is defined as the fraction of true positives and true negatives. As follows, it is calculated:



$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

Figure.6 Comparison of Accuracy (%)

Figure 6 shows that the comparison of accuracy. From the graph, it is observed that the accuracy of GFFFS-MFCM increases compared with the other plant yield prediction approaches.

5. Conclusion

In this article, a novel optimization algorithm is proposed for plant yield prediction. Initially, different plant, soil, and weather images are gathered and pre-processed for extracting the features. After extracting features, the most suitable features are selected during the classification process to reduce the search space. For feature selection process, two novel algorithms are applied such as NDSF and GF based optimization algorithm. By using these algorithms, the search space and Pareto-front issues are effectively removed and also improves the movement of fireflies during feature selection by random walk concept. The obtained best features are then classified by using MFCM scheme for predicting the plant yield. Finally, the experimental results demonstrate that the GFFFS-MFCM achieves better accuracy than the NDSFFFS-MFCM based plant yield prediction.

References:

[1] **Ramesh K and Samraj A,** Irrigation Optimization for Fruit Farms of Namakkal Area by Fuzzy Rules on Weather Parameters, J of Sci & Ind Res, 77(12) (2018) 700-704.

[2] **Choudhury, A., & Jones, J.** (2014). Crop yield prediction using time series models.*Journal of Economics and Economic Education Research*, *15*(3), 53.

[3] **Chawhan, S., Ghodichor, P., &Paunikar, K.** (2017). Survey on data mining for increasing agricultureproductivity.*International Journal for Research in Applied Science & Engineering Technology*, *5*(3), 278-280.

[4] **Cheng, H., Damerow, L., Sun, Y., &Blanke, M**. (2017). Early Yield Prediction Using Image Analysis of Apple Fruit and Tree Canopy Features with Neural Networks.*Journal of Imaging*, *3*(1), 6.

[5] **Papageorgiou, E. I., Aggelopoulou, K. D., Gemtos, T. A., & Nanos, G. D.** (2013). Yield prediction in apples using Fuzzy Cognitive Map learning approach. *Computers and electronics in agriculture*, *91*, 19-29.

[6] **Bornn, L., &Zidek, J. V**. (2012). Efficient stabilization of crop yield prediction in the Canadian Prairies.*Agricultural and forest meteorology*, *152*, 223-232.

[7] **Tsai, C. W., Huang, Y. T., & Chiang, M. C. (2014).** A non-dominated sorting firefly algorithm for multi-objective optimization. In *Intelligent Systems Design and Applications (ISDA), 2014 14th International Conference on* (pp. 62-67). IEEE.

[8] **Farahani, S. M., Abshouri, A. A., Nasiri, B., &Meybodi, M.** (2011).A Gaussian firefly algorithm.*International Journal of Machine Learning and Computing*, *1*(5), 448.

[9] **Papageorgiou, E., Markinos, A., &Gemtos, T.** (2010). Soft computing technique of fuzzy cognitive maps to connect yield defining parameters with yield in cotton crop production in central Greece as a basis for a decision support system for precision agriculture application. *Fuzzy Cognitive Maps*, 325-362

[10] **K.Ramesh, Andrews samraj.:** Identification of weed growth and intrusion in plant beds by modified Singular Value Decomposition of areal sensory images. IEEE Conference on emerging Devices and smart System., 2017;26-31

[11] **K.Ramesh, Andrews samraj.:** Weed growth and intrusion detection by multi angle images of vegetable plants using modified SVD. IEEE Conference on information technology. 2017;351-355

[12] **D. Sabareeswaran**, **Dr. R. Guna Sundari:** A Hybrid of Plant Leaf Disease and Soil Moisture Prediction in Agriculture Using Data Mining Techniques. Volume 12, Number 18 (2017) pp. 7169-7175

[13] **D.Sabareeswaran, Dr.R.Guna sundari:** Plant Yield Prediction Model Using Firefly based Feature Selection with Modified Fuzzy Cognitive Maps, Volume 118 No. 5 2018, 339-352