wjert, 2021, Vol. 7, Issue 5, 180-192. Research Article ISSN 2454-695X

World Journal of Engineering Research and Technology

WJERT

www.wjert.org SJIF Impact Factor: 5.924

MULTI-MODEL ENSEMBLE-MODIFIED DEPTH ADAPTIVE DEEP NEURAL NETWORK FOR CROP YIELD PREDICTION

Dr. M. Saranya*[1], Dr. S. Sathappan[2]**

[1]Assistant Professor, Department of Computer Science, Nandha Arts and Science College, Erode, Bharathiar University, Tamil Nadu, India.

[2]Associate Professor (Retd), Department of Computer Science, Erode Arts and Science College, Erode, Bharathiar University, Tamil Nadu, India.

Article Received on 29/06/2021 Article Revised on 19/07/2021 Article Accepted on 09/08/2021

***Corresponding Author**
Dr. M. Saranya
Assistant Professor, Department of Computer Science, Nandha Arts and Science College, Erode, Bharathiar University, Tamil Nadu, India.

ABSTRACT

Our recent study using crop yield prediction associated with climate, weather and soil for crop yield prediction. By combining regression models with Neural Networks (NN), can able to release highly satisfactory forecasting of crop yield. Prediction of crop yield accurately for tracking crop production is a trendy issue and it is a main area of research for agriculture studies. Multi-Model Ensemble Modified Depth Adaptive Deep Neural Network (MME-MDADNN) was an effective crop yield prediction method where the variation of climate, weather and soil parameters were learned through DNN. The existing Support Vector Regression (SVR) model with Deep Neural Network, the slow convergence speed, possibility of stuck in local minima and risk of over-fitting problems are resolved. Here the Ridge Regression (RR) model was applied to analyze multicollinearity in multiple regression data which enrich the better solution. Hence by applying ridge regression along with DNN, the crop yield prediction Accuracy is improved. The effectiveness of the proposed MME-MDADNN is tested in terms of Accuracy, Precision, Recall and F-measure.

KEYWORDS: Multiple Linear Regression, SVR, Ridge Regression, Neural Networks, Yield Prediction.

I. INTRODUCTION

Agriculture is the main source of cultivating the soil, growing crops and raising livestock. It provides most of the world's food and fabrics, Cotton, wool, and leather are all agricultural products and wood for construction and paper products.[1] In the world's surface, around 71% covered by water, which contributes only 2% of human food. 20% of the land area is suitable for agriculture. So only the 6% of world's surface must produce the food for human's need. The human population became more than doubled in the last 50 years from 2.5 billion to 10.5 billion today and will reach almost 20 billion people in the next 50 years.[2] In current era, the large numbers of farmers are not getting the planned profit because of many challenges in the climate, weather and seasonal changes. They need timely guidance for their potential profit about their crops.[3] So an analysis is made to gain and increase the profit. During, the last decade, prediction of crop yield carried out via manually, by analyzing cultivator's previous experiences on the specified crop.[4]

In this paper, the MME-MDADNN is proposed for modeling climate changes, soil parameters and climate parameters such as kprecipitation, hot, cool, normal temperature, cloud density, vapor pressure, wet day, dry day frequency and humidity to predict the crop yield. To enhance the Accuracy of crop yield prediction by considering multiple parameters related to climate. Based on the historical data about significant phenomena and the operations concerning to the result, the MME-MDADNN enhances the Accuracy of standard DNN. The Accuracy of MME-MDADNN is improved by considering multiple parameters of climate and the effect of weather on crop yield for crop yield prediction.

The rest of the article is prepared as follows: Section II studies the researches related to the crop yield prediction. Section III describes the functioning of MME-MDADNN for crop yield prediction and Section IV portrays its performance. Section V discusses the conclusion about this research work.

II. Literature Survey

In this section, the discussion of various new recent techniques used to predict crop yield productivity in various regression techniques. The advantages and disadvantages also discussed in this section.

Van et. al.[5] proved that all the farmers are interested to know about current changes in the crop yield. Hence, the amount of agricultural information is huge and analysis made by manual is very difficult. The machine learning is used as the field of science, allows it made easy to learn machinery without specific programming. It is emerged that, agricultural applications to enhance the predictive Accuracy of the crop yield. Although machine learning[6] is considerably improved, it is used in data driven ways is limited. Also, the Accuracy depends on the model representativeness, data quality and the reliance between the target and input variables in the collected dataset.

Khaki et. al.[7] proved that deep learning techniques are used to enhance the Accuracy of predicting the crop yields. The DNN for predicting the crop yield depicts the actual information with no handcrafted characteristics. DNN was used to model nonlinear and complicated relationship among input parameters and it applied their prior details and achieved a precise prediction of yield from known weather conditions. However, this technique required more advanced model to learn impact of various parameters and their changes.

Aghighi et. al.[8] designed several machine learning techniques, namely Random Forest Regression (RFR), Boosted Regression Tree (BRT), Gaussian Process Regression (GPR) and Support Vector Regression (SVR) and approaches to analyze the silage maize yield prediction. The experiment results were takem on time series of Normalized Difference Vegetation Index (NDVI) dataset, which was derived from Landsat 8 OLI images. The results proved that the BRT, GPR and RFR achieved higher performance compared to conventional regression methods because those techniques handle with high-dimensional data of complex distributions as well as the inconsistency of NDVI time series. Hence, this approach did not handle the effect of other environmental parameters such as soil moisture, texture of soil, climatic parameters.

E. Khosla, R. Dharavath, and R. Priya,[9] developed a Modular Artificial Neural Network (MANN) for analyzing the rainfall amount to improve the yield of major kharif crops in Visakhapatnam. The important features were selected by using SVR to identify the various types of kharif crops, bajra, rice, ragi and maize. The experiments were done on collected data of recent year and compared it with crop data of the year 1997. The MANN-SVR method used only on rainfall and area attribute in predicting the yield of crops, but the yield of the certain crop depends on many other factors like fertilizers used, irrigation and many more.

P. M. Gopal et. al.[10] proposed the hybrid Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) to identify the accurate crop yield. MLR intercept and coefficients were applied to initialize the ANN's input layer bias and weights. The results proved that the prediction Accuracy was increased and obtained less number of optimal errors. The prediction Accuracy was calculated by using standard performance metrics and the results were compared with traditional techniques including conventional statistical model MLR, conventional ANN, SVR, KNN and random forest. The results signed that the proposed hybrid MLRANN model gave much better prediction Accuracy than other models of the same agricultural dataset. The production value decrease, when the maximum temperature increased beyond the threshold value.

III Proposed methodology

There is emerge need to promote the technical ability of the farmers to make them competent with the current challenging technological developments happening around the world. The success of agricultural production mainly depends on selecting the appropriate crop suitable for the agricultural land.[11]

In this section, the Multi-Model Ensemble with Modified Depth Adaptive Deep Neural Network (MME-MDADNN) is described in detail for crop yield prediction. At First, the climate, weather and soil related at a particular area are collected and it is pre-processed with multiple imputation techniques. A statistical model is processed on the pre-processed data to predict the variation of data from year-to-year. This information is given as input to the MDADNN for crop yield. The MDADNN is the combination of DADNN and RR where RR is used in the top most layer of DADNN.

The overall architecture of DADNN is depicted in Figure 3.1.

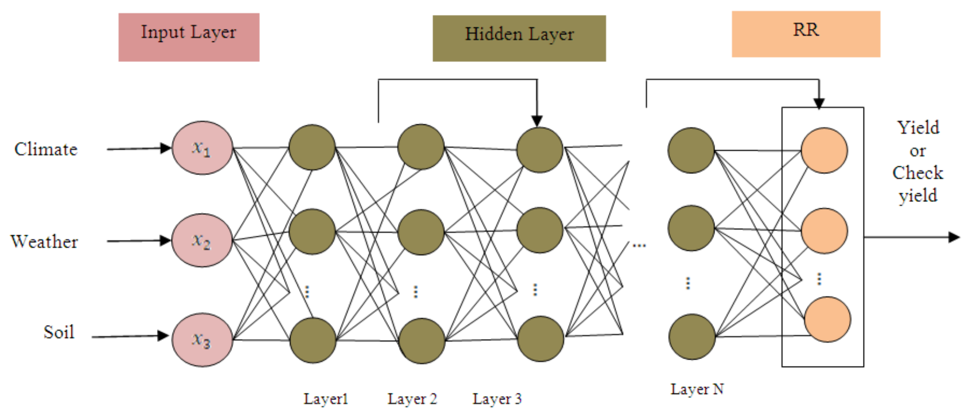

Figure 3.1 Architecture of Modified Depth Adaptive Deep Neural Network.

3.1.1 Problems involved in Depth Adaptive Deep Neural Network

In MME-EDADNN[12], the input layer and hidden layer are used for feature extraction. The output layer is used for the feature extraction using the softmax and cross entropy for crop yield prediction. The output layer with softmax and cross entropy is known as multinomial LR. The multinomial LR is more sensitive to outliers, does not go to zero even if the point is classified sufficiently confidently and also might lead to minor degradation in prediction Accuracy. So, the SVR is used to rectify the problems in the MME-DADNN.

In MME-DADNN[13], forward or backward selection could not be able to tell anything about the removed variables effect on the response. Removing predictors from the model can be seen as settings their coefficients to zero. Instead of forcing them to be exactly zero, let it to penalize them if they are too far from zero, thus enforcing them to be small in a continuous way. This way is used to decrease model complexity while keeping all variables in the model. This basically, is what Multi Model Ensemble Modified Depth Adaptive Deep Neural Network does.

3.1.2 Modified Depth Adaptive Deep Neural Network

In order to solve the problems in DADNN, the MDADNN is constructed for efficient crop yield prediction. The MDADNN is the combination of DADNN and RR. The RR is included in the top most layer of DADNN that replaces the multinomial LR in MME-DADNN. For the given input x_i (i.e., soil, weather and climate parameters), the prediction function of MME-MDADNN is given as follows:

$$Y = XD + r \qquad (1)$$

Where Y is the dependent variable, X represents the independent variables, D is the regression coefficients to be estimated, and r represents the errors are residuals.

METHODOLOGY

Here, the Ridge Regression, minimize the sum of squared residuals but also penalize the size of parameter estimates, in order to shrink them towards zero:

$$L_{ridge}(\alpha)=\sum(y_i-x_i\alpha)^2+\delta\sum\alpha\|y_j-X\,\alpha\|^2+\delta\|\,\alpha\|^2 \quad (2)$$

In Eq(2), i=1,2…n and j=1,2,…m

The ridge regression estimates $\alpha_{ridge}=(X'X+\delta I)^{-1}(X'Y)$,

where *I* denotes the identity matrix.

The δ parameter is the regularization penalty.

- As $\delta \rightarrow 0$, $\alpha_{ridge} \rightarrow \alpha_{ridge}$ Ordinary Least Square(OLS);
- As $\delta \rightarrow \infty$, $\alpha_{ridge} \rightarrow 0$;

So, setting δ to 0 is the same as using the OLS, while the larger its value, the stronger is the coefficients' size penalized.

Bias-Variance in Ridge Regression

Incorporating the regularization coefficient in the formulae for bias and variance gives us

$\mathrm{Bias}(\alpha_{ridge})=-\delta((X'X+\delta I)^{-1}(X'Y)\alpha$, $\mathrm{Var}(\alpha_{ridge})=\Omega^2(X'X+\delta I)^{-1}X'X(X'X+\delta I)^{-1}$ (3)

Minimizing Information Criteria

In this method degrades the estimating the model with many different values for λ and choosing the one that minimizes the Bayesian Information Criterion:

$AIC_{\ ridge}= n \ln (e'e)+2df_{ridge}$ (4)

$BIC_{ridge}=n \ln(e'e)+2df_{ridge}\ln(n)$ (5)

where df_{ridge} is the number of degrees of freedom. The degrees of freedom are equal to the trace of the so-called *hat matrix*, which is a matrix that maps the vector of response values to the vector of fitted values as follows: y^=Hy.

In ridge regression, however, the formula for the hat matrix should include the regularization penalty: $H_{ridge} = X(X'X + \delta I)^{-1}X$, which gives $df_{ridge} = trH_{ridge}$, which is no longer equal to m.

Minimizing cross-validated residuals

To choose δ through cross-validation, choose a set of R values of δ to test, split the dataset into m folds, and follow this algorithm:

for r in 1:R:

for m in 1:M:

keep fold m as hold-out data

use the remaining folds and $\delta = \delta_p$ to estimate α^ridge

predict hold-out data: x_{test}, m=X_{test}, m(α^ridge)

compute a sum of squared residuals: $SSR_k = ||x - x_{test,m}||^2$

end for m

average SSR over the folds: SSR_R=1K∑m_m=1SSR_m

end for r

choose optimal value:

$\delta_{opt} = argmin_r SSR_r$

4. RESULT AND DISCUSSION

Here, the performance of MME-DADNN, MME-EDADNN and MME-MDADNN is evaluated in terms of Accuracy, Precision, Recall and F-measure. For the experimental purpose,

The crop data are collected from https://data.world/thatzprem/agriculture-india

The climate data are collected from http://www.ccafs-climate.org/climate wizard /

The soil data are collected from https://data.gov.in/search/site?query=soil

From the collected data, 70,000 data are used for training process and 30,000 data are used for testing process.

4.1 Accuracy

Accuracy is the fraction of true positive and true negative among the total number of samples examined. It is calculated as incipient,

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

TP: True Positive

TN: True Negative

FP: False Positive

FN: False Negative

Table 4.1: Evaluation of Accuracy.

Crop yield prediction method	Banana	Groundnut	Wheat	Sugarcane	Maize
MME-DADNN	0.94	0.95	0.94	0.95	0.94
MME-EDADNN	0.97	0.965	0.96	0.97	0.97
MME-MDADNN	0.98	0.97	0.97	0.98	0.98

Table 4.1 tabulates the Accuracy value of MME-DADNN, MME-EDADNN and MME-MDADNN methods for crop yield prediction.

The Accuracy of MME-EDADNN is 3.09%, 1.55%, 2.08%, 2.06% and 3.09%, MME-MDADNN is 4.08%, 2.06%, 3.09%, 3.06% and 4.08% greater than MME-DADNN method for banana, groundnut, wheat, sugarcane and maize respectively. In this analysis, it is proved that the proposed MME-MDADNN has high Accuracy for five crops than MME-EDADNN and MME-DADNN based crop yield prediction method.

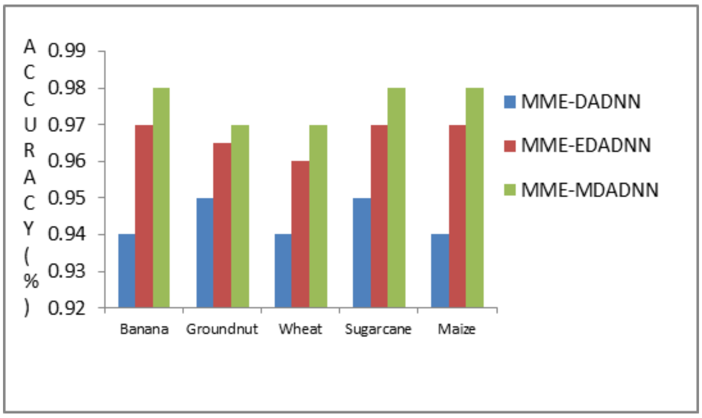

Figure 4.1 Evaluation of Accuracy.

Figure 4.1 and Table 4.1 present the Accuracy of MME-DADNN, MME-EDADNN and MME-MDADNN for five different crops.

4.2 Precision

The value of Precision is calculated based on the crop yield prediction at TP and FP rates. It is calculated as,

$$Precision = \frac{TP}{TP + FP}$$

Table 4.2 tabulates the Precision value of MME-DADNN, MME-EDADNN and MME-MDADNN methods for crop yield prediction.

Table 4.2: Evaluation of Precision.

Crop yield prediction method	Banana	Groundnut	Wheat	Sugarcane	Maize
MME-DADNN	0.62	0.934	0.89	0.9	0.65
MME-EDADNN	0.75	0.96	0.93	0.945	0.78
MME-MDADNN	0.76	0.97	0.95	0.95	0.80

Table 4.2 and Figure 4.2 show the performance comparison of the proposed and existing crop yield methods in terms of Precision rate. The performance is evaluated using banana, groundnut, wheat, sugarcane and maize crops. The crops are represented in x-axis whereas Precision rate is represented in y-axis.

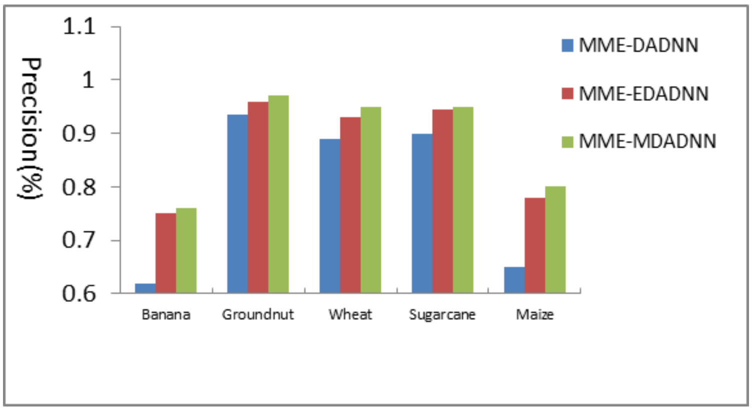

Figure 4.2 Evaluation of Precision.

MME-MDADNN has 18.42%, MME-EDADNN has 17.33% increased Precision value than MME-DADNN for banana crop. While the value of MME-MDADNN is 3.71% and MME-EDADNN is 2.71% greater than MME-DADNN for groundnut crops, Wheat has 4.30% and 6.32% increase than existing model and sugarcane records 4.76% and 5.26% increase in MME-EDADNN and MME-MDADNN and also maize is with 16.67% and 18.75% higher Precision value than the value of MME-DADNN and MME-EDADNN respectively. Figure 4.2 and Table 4.2 prove that the proposed MME-MDADNN has achieved high Precision than MME-EDADNN and MME-DADNN based crop yield prediction method.

4.3 Recall

The value of Recall is calculated based on the crop yield prediction at TP and FN rates.

$$Recall = \frac{TP}{TP + FN}$$

Table 4.3 tabulates the Recall value of MME-DADNN, MME-EDADNN and MME-MDADNN methods for crop yield prediction.

Table 4.3: Evaluation of Recall.

Crop yield prediction method	Banana	Groundnut	Wheat	Sugarcane	Maize
MME-DADNN	0.95	0.94	0.945	0.95	0.942
MME-EDADNN	0.972	0.96	0.966	0.975	0.964
MME-MDADNN	0.98	0.97	0.98	0.978	0.971

From the Table 4.3, the Recall of MME-MDADNN for banana crop is 2.26% and 0.82% greater than MME-EDADNN and MME-DADNN respectively. While groundnut and wheat crops have 2.08%, 1.03% and 2.17%, 0.31% more value than MME-EDADNN and MME-

DADNN respectively. Sugarcane and maize have 2.56%, 0.31% and 2.28%, 0.72% increased value than MME-EDADNN and MME-DADNN.

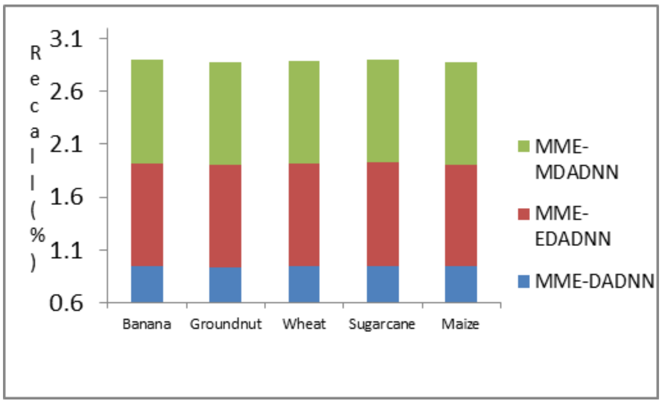

Figure 4.3 Evaluation of Recall.

As the values in Figure 4.3 shows the Recall of MME-MDADNN, MME-EDADNN and MME-DADNN for five different crops. Banana, groundnut, wheat, sugarcane and maize crops are taken in x-axis and the y-axis shows Recall. Thus the proposed MME-MDADNN has high Recall than the MME-EDADNN and MME-DADNN based crop yield prediction method.

4.4 F-measure

F-measure is calculated by using the values of both Precision and Recall. It is calculated as,

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

Table 4.4 tabulates the F-measure value of MME-DADNN, MME-EDADNN and MME-MDADNN methods for crop yield prediction.

Table 4.4: Evaluation of F-measure.

Crop yield prediction method	Banana	Groundnut	Wheat	Sugarcane	Maize
MME-DADNN	0.95	0.94	0.945	0.95	0.942
MME-EDADNN	0.975	0.962	0.978	0.976	0.97
MME-MDADNN	0.98	0.97	0.98	0.98	0.99

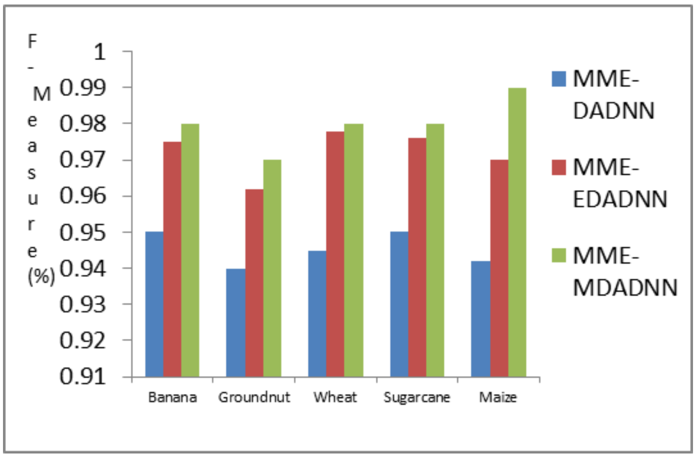

Figure 4.4 Evaluation of F-measure.

The F-measure of MME-MDADNN is 2.56% greater than MME-DADNN and 0.51% greater than MME-EDADNN for banana crop while Groundnut has 2.29% and 0.82% higher value than MME-EDADNN and MME-DADNN. Wheat gives 3.37% and 0.20% increased value than the existing models. Sugarcane records 2.66% and 0.41% greater F-measure value than MME-EDADNN and MME-DADNN whereas 2.89% greater value is observed than MME-DADNN and 2.02% higher value is observed than MME-EDADNN for maize crop.

From Table 4.4, the result shows the proposed MME-MDADNN has got high F-measure when compared to MME-EDADNN and MME-DADNN based crop yield prediction methods. The comparison of MME-DADNN, MME-EDADNN and MME-MDADNN in terms of F-measure value is shown in Figure 4.4. The performance comparison of proposed and existing crop yield prediction method is evaluated using five different crops. x-axis has collected crop values and y-axis has F-measure values.

5. CONCLUSION

In this paper, a MME-MDADNN method is proposed for efficient crop yield prediction. Initially, the variation of soil, weather and climate parameters over a time period and is analyzed using statistical model and it is given as input to DNN for soil, weather and climate predictions. The predicted soil, weather and climate parameters are processed in DADNN which extract the features using the input and hidden layers. The extracted features are processed in SVR to predict the crop yield. By using SVR in DADNN, the slow convergence speed, possibility of stuck in local minima and risk of over-fitting problems are resolved. The MME-EDADNN is compared with MME-MDADNN by applying ridge regression method instead of SVR. Hence the experimental results prove that the proposed MME-MDADNN method has high Accuracy, Precision, Recall and F-measure than MME-DADNN and MME-EDADNN method for crop yield prediction.

REFERENCES

1. Dubock AC, "Biofortified Golden Rice as additional intervention for vitamin A deficiency", Cambridge: Burleigh Dodds Science Publishing Limited, 2017; 32(11): 119–220.
2. Kyndta T, Quispea D, Zhaic H, Jarret R, "The genome of cultivated sweet potato contains Agrobacterium T-DNAs with expressed genes: an example of a naturally transgenic food crop", PNAS, 2015; 112(18): 5844–49.
3. Kalaiarasi E, Anbarasi A, "Crop yield prediction using multi-parametric deep neural networks", Indian Journal of Science and Technology, 2021; 14(2): 131-140.
4. Khaki S, Wang L, Archontoulis SV, "A CNN-RNN Framework for Crop Yield Prediction", Frontiers in Plant Science, 2019; 10: 01750.
5. Van Klompenburg T, Kassahun A, Catal C, "Crop yield prediction using machine learning: A systematic literature review", Computers and Electronics in Agriculture, 2020; 177(10): 105709.
6. Champaneri M, Chachpara D, Chandvidkar C, Rathod M, "Crop yield prediction using machine learning", International Journal of Science and Research, 2020; 4: 1–4.
7. Khaki S, Wang L, "Crop Yield Prediction Using Deep Neural Networks", Frontiers in Plant Science, 2019; 10: 1–10.
8. Aghighi H, Azadbakht M, Ashourloo D, Shahrabi H.S, and Radiom S, "Machine Learning Regression Techniques for the Silage Maize Yield Prediction Using Time-Series Images of Landsat 8 OLI", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018; 11(12): 4563-4577.
9. Khosla E, Dharavath R, and Priya R, "Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression", Environment, Development and Sustainability, 2019; 1-22.
10. Gopal P.M, and Bhargavi R, "A novel approach for efficient crop yield prediction", Computers and Electronics in Agriculture, 2019; 165: 104968.
11. Khamparia A, Saini G, Gupta D, Khanna A, Tiwari S, "Seasonal Crops Disease Prediction and Classification Using Deep Convolutional Encoder Network", Circuits, Systems, and Signal Processing, 2019; 1-19.

12. Saranya, M., & Sathappan, S. Multi-model ensemble-enhanced depth adaptive deep neural network for the crop yield prediction, ICMLDA, 2019.
13. Saranya, M., & Sathappan, S. Multi-model ensemble depth adaptive deep neural network for crop yield prediction. International Journal of Recent Technology and Engineering (IJRTE), 2019; 8(5): 3088-3093.