

RESEARCH ARTICLE



Crop yield prediction using multi-parametric deep neural networks

E Kalaiarasi^{1*}, A Anbarasi²

¹ Ph.D. Research Scholar, Department of Computer Science, L. R. G. Government Arts College for Women, Tiruppur, Tamilnadu, India

² Assistant Professor, Department of Computer Science, L. R. G. Government Arts College for Women, Tiruppur, Tamilnadu, India

 OPEN ACCESS

Received: 23.11.2020

Accepted: 28.12.2020

Published: 20.01.2021

Citation: Kalaiarasi E, Anbarasi A (2021) Crop yield prediction using multi-parametric deep neural networks. Indian Journal of Science and Technology 14(2): 131-140. <http://doi.org/10.17485/IJST/v14i2.2115>

* **Corresponding author.**

kalaiarasiephd@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2021 Kalaiarasi & Anbarasi. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.isee.org/))

ISSN

Print: 0974-6846

Electronic: 0974-5645

Abstract

Objective: To propose Multi-parametric Deep Neural Network (MDNN) for modeling the impact of climate changes, multiple parameters related to the weather and soil for accurate crop yield prediction. **Methods:** In MDNN, a measure called Growing-Degree Day (GDD) is introduced for measuring the overall effect of weather conditions related to the crop yield. One of the key elements in MDNN is the neuron's layer-wise activation function. In order to enhance the crop yield predictive performance, a leaky rectified linear unit is used in the activation units of MDNN. For the analysis of performance of DNN and MDNN, data about weather, crop and soil are collected from <http://www.ccafs-climate.org/climatewizard/>, <https://data.world/thatzprem/agriculture-india> and <https://data.gov.in/search/site?query=soil> respectively. From the collected data, 60000 records are used for training and 40,000 records are used for testing. **Findings:** By considering multiple parameters of climate and the effect of weather on crop yield, the accuracy of MDNN is improved for predicting the crop yield. The effectiveness of MDNN is tested and compared with DNN for different types of crops. The MDNN achieves 91.84% of mean accuracy for five different crops compared to the DNN classification. **Novelty:** This proposed work tries to predict the crop yield more accurately by analyzing the climate, weather and soil parameters. The MDNN considerably improves statistical efficiency over typical DNN by using previous knowledge about important phenomena and functional forms relating them to the crop yield.

Keywords: Crop yield prediction; machine learning; DNN; climatic changes; soil parameters; growing degree-day

1 Introduction

Roughly one third of the land on Earth is used for farming, with the majority of it being used for agriculture. Agricultural demand is increasing due to the growing world population. The control of farming activities is faced by unique problems including the seasonal cycles of crop phenology and crop production depends on weather, climate and soil parameters. A large number of farmers don't get the planned profit because of these challenges. Farmers require timely guidance on the potential profitability of crops,

so an analysis must be conducted to assist farmers to increase their profit. One of the major issues in agriculture field is yield prediction^(1,2).

Each and every farmer is interested in knowing about the crop yield. During previous centuries, prediction of crop yield has been carried out via manually analyzing the cultivator's prior knowledge on the specified crop. However, an amount of agricultural information is high and manual analysis is quite difficult. Machine learning⁽³⁾ is defined as the field of science, which allows it possible to learn machinery without specific programming. Nowadays, machine learning⁽⁴⁾ is emerged in agricultural applications to boost the predictive accuracy of the crop yield and resolve the problem in manual analysis. Although machine learning⁽⁵⁾ is considerably improved, its use in data driven ways is limited. Also, its accuracy depends on the model representativeness, data quality and the reliance between the target and input variables in the collected dataset.

So, deep learning techniques have been used to enhance the accuracy of predicting the crop yields. The DNN for predicting the crop yield⁽⁶⁾ represents the actual information with no handcrafted characteristics. DNN was designed to train nonlinear and complicated relationship among input parameters using their prior details and achieve a precise prediction of yield from known weather conditions. However, this method required more advanced model to learn impact of varied parameters and their changes.

In this paper, the MDNN is proposed for modeling climate changes, soil parameters and climate parameters such as precipitation, temperature, cloud cover, vapor pressure, wet day frequency and humidity to predict the crop yield. It enhances the accurateness of predicting the crop yield by considering multiple parameters related to climate. Based on the historical data about significant phenomena and the operations concerning to the result, the MDNN enhances the accuracy of standard DNN. Through extending a parametric framework with DNN, the dynamics are captured which are either missing or improperly defined in the parametric frameworks. The accuracy of MDNN 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 MDNN for crop yield prediction and Section IV portrays its performance. Section V summarizes this research work and suggests the future scope.

2 Literature Survey

A hybrid approach⁽⁷⁾ was proposed using Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) for crop yield prediction. However, in hybrid approach input weights reach near global minima at an early stage of iteration which leads to premature convergence. A neural collaborative filtering approach⁽⁸⁾ was presented for predicting yield performance of parents in plant breeding. The prediction performance of this approach will be enhanced by including other important parameters such as weather components and soil conditions.

Deep learning approaches⁽⁹⁾ were introduced for winter wheat yield prediction in main wheat-producing regions of china. However, the yield de-trending could possibly cause high uncertainty of yield prediction and should be handled carefully when applied in crop yield prediction.

An integrated model⁽¹⁰⁾ was proposed for rice yield prediction with respect to climate change. In this model, the changes in the yield were attributed to crop quality and climate parameters, and the regularized variance vegetation value from MODIS was applied to forecast the crop condition. Because of the significance, the input parameters were minimized by the spatial interpolation. It has more sophisticated issues such as changes in CO₂ level and soil characteristics when it comes to forecasting yield for the far future.

A prediction model⁽¹¹⁾ was introduced for crop yield prediction method for dataset collected from various region of Bangladesh. This method utilized various soil attributes for crop yield prediction. The climatic and soil variables were processed in Density-Based Spatial Clustering of Applications with Noise (DBSCAN), K-means, Clustering Large application (CLARA) and Partition Around Medoids (PAM) to cluster the data. After that, linear regression method was applied on the clustered data for crop yield prediction. But, this model was only suitable for small datasets.

Artificial Neural Network (ANN)⁽¹²⁾ has been suggested to predict maize yield in South Africa. It was processed through the climate variables such as maximum temperature, potential evapotranspiration, minimum temperature, soil moisture, precipitation and land cultivated for maize. These variables were processed in the input, hidden and output layer of ANN for prediction of maize yield. But, the difficulty of this model was increased while increasing the amount of neurons used in ANN.

Paddy yield predictor⁽¹³⁾ was proposed for paddy yield prediction with respect to nitrogen, temperature, rainfall, and soil pH. This predictor makes use of agricultural data for finding the correlation and paddy yield prediction. Association rule mining techniques including Apriori, AprioriTid and Eclat were utilized in this predictor for finding the relation between paddy yield, soil nitrogen, rainfall, and soil pH for paddy yield prediction. This work will be extended for predicting the crop type that can

be cultivated in a particular plot for high yield.

3 Proposed Methodology

In this section, the MDNN for predicting the crop yields is explained briefly.

3.1 Neural network based weather prediction

Initially, the agriculture data are collected and the weather is predicted using neural network. Consider the weather attribute $V_{l,t}^w$ at locality l in the past 4 years t as $\{X_{l,t-1}^w, X_{l,t-2}^w, X_{l,t-3}^w, X_{l,t-4}^w\}$. The weather attributes are denoted as $f(x)$ which is given as input to the neural network. The hidden layer of DNN is defined as tan-sigmoid transfer function.

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{1}$$

In Eq. (1), x is the climate attribute. Each climate attribute has its own weight values as w_1, w_2, \dots, w_n and the weighted sum of the inputs is done by the adder function as:

$$u = \sum_{i=1}^n w_i x_i \tag{2}$$

In Eq. (2), n stands for the amount of attributes. The output layer of DNN is described by:

$$y_t^w = f(\sum_{i=1}^n w_i x_i + b_i) \tag{3}$$

In Eq. (3), y_t^w is the weather attributes at time t , $f(x)$ is the transfer function, w_i is the weight values, x_i is the weather attributes and b_i refers to the bias value. Figure 1 shows the basic structure of DNN for weather prediction.

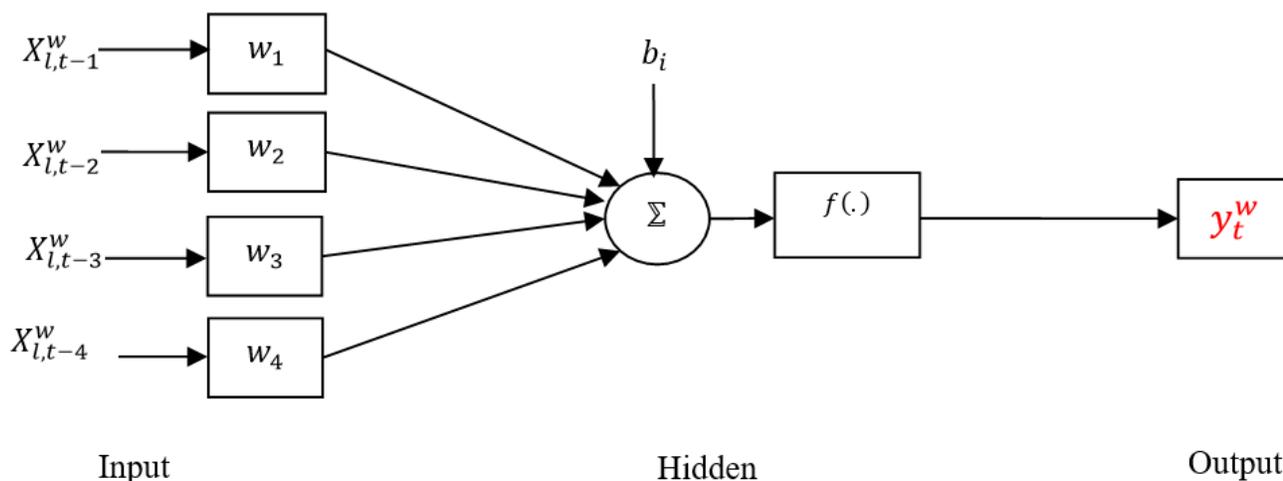


Fig 1. Fundamental design of DNN for weather prediction

Similarly, climate attributes $V_{l,t}^c$ at locality l in the past 4 years t is denoted as $\{X_{l,t-1}^c, X_{l,t-2}^c, X_{l,t-3}^c, X_{l,t-4}^c\}$. By processing the climate attributes in DNN, the climate attributes at time $t y_t^c$ is predicted. The soil attributes $V_{l,t}^s$ at locality l in the past 4 years t is denoted as $\{X_{l,t-1}^s, X_{l,t-2}^s, X_{l,t-3}^s, X_{l,t-4}^s\}$. By processing the climate attributes in DNN, the soil attributes at time $t y_t^s$ is predicted.

3.2 Impact of climate changes to crop yield

The growth of crops highly depends on the climate attributes such as precipitation, temperature, cloud cover, vapor pressure, wet day frequency and humidity. The effect of multiple climate attributes on the growth of crops is measured using GDD. An Ordinary Least Square (OLS) regression¹³ is used in GDD as:

$$y_{it} = \alpha_i + \sum_r GDD_{rit} + X_{it}VPD_{it}\beta + \varepsilon_{it} \tag{4}$$

In Eq. (4), y_{it} is the impact of climate attributes on crop yield i and t , α_i is the intercept of the model, X_{it} denotes the climate attributes on crop i and t , r denotes the value of every GDD bin, X represents all climate attributes, β is the dependent variable to and ε is the error term. The amount of water in air influence the climate that also affects the crop yield which is calculated using Vapor Pressure Deficit (VPD). It is calculated for the standard vapor pressure at a day’s maximum temperature. VPD_{it} is calculated as,

$$VPD_{it} = 0.6107 \left(e^{\left(\frac{17.269T_{h_it}}{237.3 + T_{h_it}} \right)} - e^{\left(\frac{17.269T_{l_it}}{237.3 + T_{l_it}} \right)} \right) \tag{5}$$

In Eq. (5), T_{h_it} and T_{l_it} denote the maximum and minimum temperature for crop i at time t . The impact of climate attributes on crop yield is calculated using Eq. (4).

3.3 Prediction of multi-parametric DNN

The major unit in MDNN is the neuron’s layer-wise activation function i.e., the Rectified Linear Unit (ReLU). But, it has a limitation on predicting the crop yield while the gradient is null i.e., during idle state. It is not able to fine-tune the weights and also reduces the learning speed due to vanishing gradient problems. To tackle these problems, a leaky ReLU (lReLU)^(14,15) is used in the activation function which permits non-zero gradient during idle state. This activation function $a(x)$ plays the role for system nonlinearity as:

$$a(x) = \begin{cases} x, & \text{if } x > 0 \\ \frac{x}{100}, & \text{if } x < 0 \end{cases} \tag{6}$$

This is a variant of ReLU: $a(x) = \max(0, x)$. Also, the output of MDNN is given as:

$$Yield = \alpha_i + X\beta + V_i^l\Gamma^l + \varepsilon_{it} \tag{7}$$

In Eq. (7), X denotes the input (y_t^c, y_t^w, y_t^s and y_{it}) (effect of weather conditions on crop yield, multiple attributes of weather and soil).

$$\begin{aligned} V_{it}^l &= a(\gamma^2 + V_{it}^2\Gamma^2) \\ V_{it}^2 &= a(\gamma^3 + V_{it}^3\Gamma^3) \\ &\vdots \\ V_{it}^{21} &= a(\gamma^{21} + X_{it}\Gamma^{21}) \end{aligned} \tag{8}$$

In Eq. (8), V^l is determined at l -th layer. The attributes Γ^l map X to the prediction of crop yield. After the prediction of yield $yield$ and yield checking y_c using DNN, the difference between yields y_d is calculated as:

$$\begin{aligned} \text{Variance}(yield_d) &= \text{Variance}(yield - y_c) \\ &= \text{Variance}(yield) + \text{Variance}(y_c) - 2Cov(yield, y_c) \end{aligned} \tag{9}$$

The yield difference is used to calculate the dissimilarity between predicted and actual yield which helps to evaluate the accuracy of DNN. Figure 2 shows MDNN design for predicting the yield of different crops.

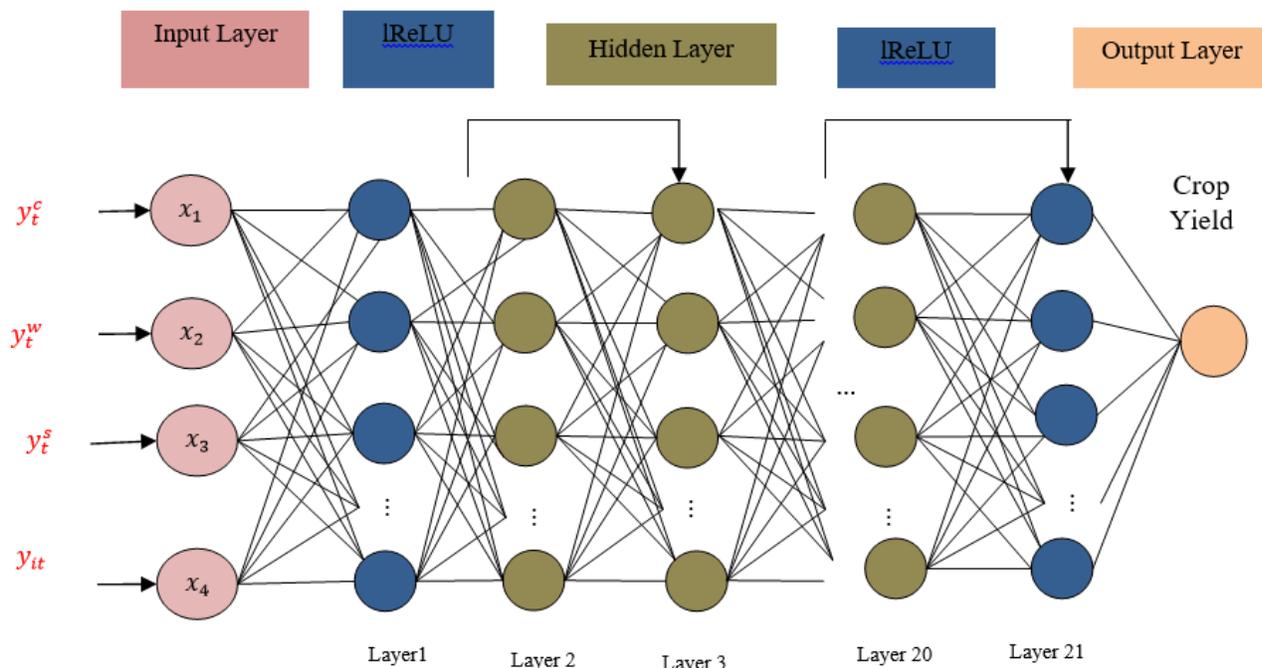


Fig 2. MDNN structure for crop yield prediction

4 Result and Discussion

In this section, the efficiency of the ⁽⁷⁾ANN, ⁽⁷⁾MLR, ⁽⁶⁾DNN and MDNN method for crop yield prediction is tested based on different metrics used for evaluating the classifier performance. In this experiment, the details of weather, crop and soil are acquired from <http://www.ccafs-climate.org/climatewizard/>, <https://data.world/thatzprem/agriculture-india> and <https://data.gov.in/search/site?query=soil>, respectively. Among the acquired data, consider 60000 are for training and 40,000 are for testing. For crop yield prediction, three different classes such as low, medium and high are identified. The classification accuracy is evaluated by comparing actual and predicted class labels of the test dataset. The parameters used in weather datasets are air temperature, humidity, humidity, soil temperature and wind. Bulk Density(BD), Soil PH, Salinity (EC), Cation Exchange Capacity(CEC), C:N ratio(Carbon to Nitrogen), Phosphates activity of Soil, Degradation level Soil Color, Soil Type & Sub Type ,Soil Texture, Soil Structure type and Soil Consistence are soil parameters used in the dataset.

The hardware on which the software is to be built is Core i4 and more with a minimum of 4GB of RAM and 500GB of hard drive. The electronic circuits to be taken into account will be added gradually, as at the stage of growth. The crop yield prediction software is written using MATLAB 2017 b.

4.1 Accuracy

It is the rate of correctly predicted crop yields over the number of predictions executed.

$$Accuracy = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{TP + TN + False\ Positive\ (FP) + False\ Negative\ (FN)}$$

Where, TP: The label of crop yield for the amount of instances high is predicted as high.

FP: The label of crop yield for the amount of instances low and medium is predicted as low and medium.

FN: The label of crop yield for the amount of instances low/medium is predicted as high and low is predicted as the medium.

FP: The label of crop yield for the amount of instances high is predicted as medium/low, the medium is predicted as low.

Table 1 represents 5 different crops yield prediction accuracy obtained for DNN and MDNN.

Table 1. Comparison of DNN and MDNN in terms of accuracy

Methods	Banana	Groundnut	Wheat	Sugarcane	Maize
ANN	0.80	0.84	0.81	0.79	0.78
MLR	0.84	0.87	0.85	0.84	0.83
DNN	0.88	0.91	0.89	0.88	0.88
MDNN	0.91	0.92	0.91	0.90	0.93

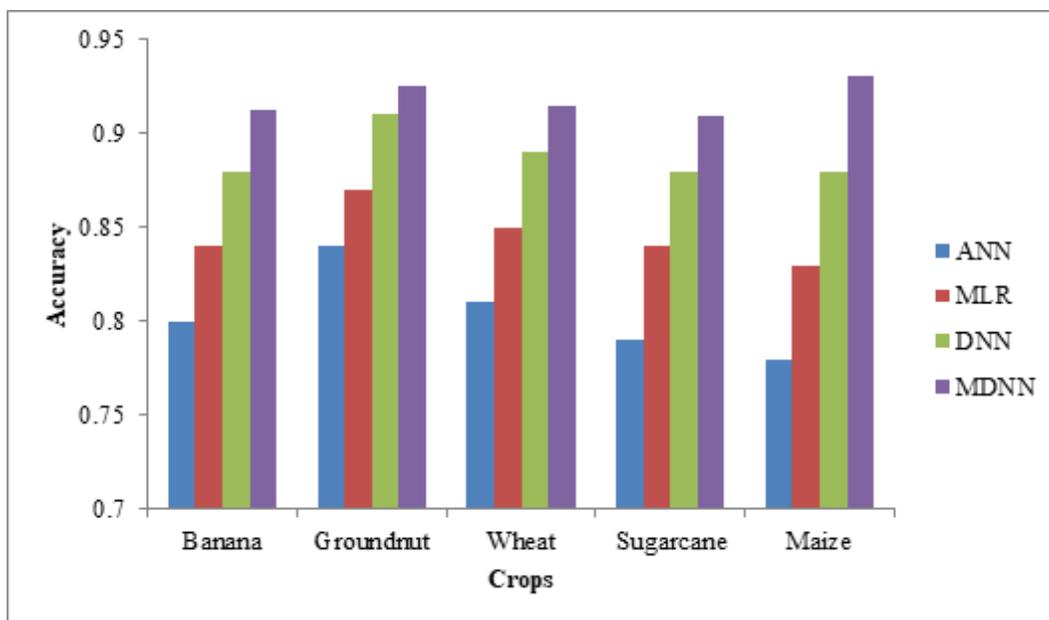


Fig 3. Comparison of accuracy for DNN and MDNN

Figure 3 show the accuracy of DNN and MDNN for predicting 5 variety crops yield. The accuracy of MDNN is 3.64%, 1.65%, 2.81%, 3.3% and 5.8% greater than DNN for banana, groundnut, wheat, sugarcane and maize crops respectively. By considering multiple parameters and the effect of weather on crop yield, the accuracy of MDNN is improved. It concludes the MDNN has a high accuracy than DNN for all considered crops yield prediction.

4.2 Precision

It is the amount of exactly predicted crop yields over TP and FP.

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

Table 2 shows the precision obtained from DNN and MDNN for predicting 5 crop yield.

Table 2. Comparison of DNN and MDNN in terms of Precision

Methods	Banana	Groundnut	Wheat	Sugarcane	Maize
ANN	0.73	0.85	0.78	0.69	0.78
MLR	0.75	0.88	0.81	0.73	0.81
DNN	0.79	0.91	0.85	0.78	0.85
MDNN	0.85	0.93	0.89	0.83	0.87

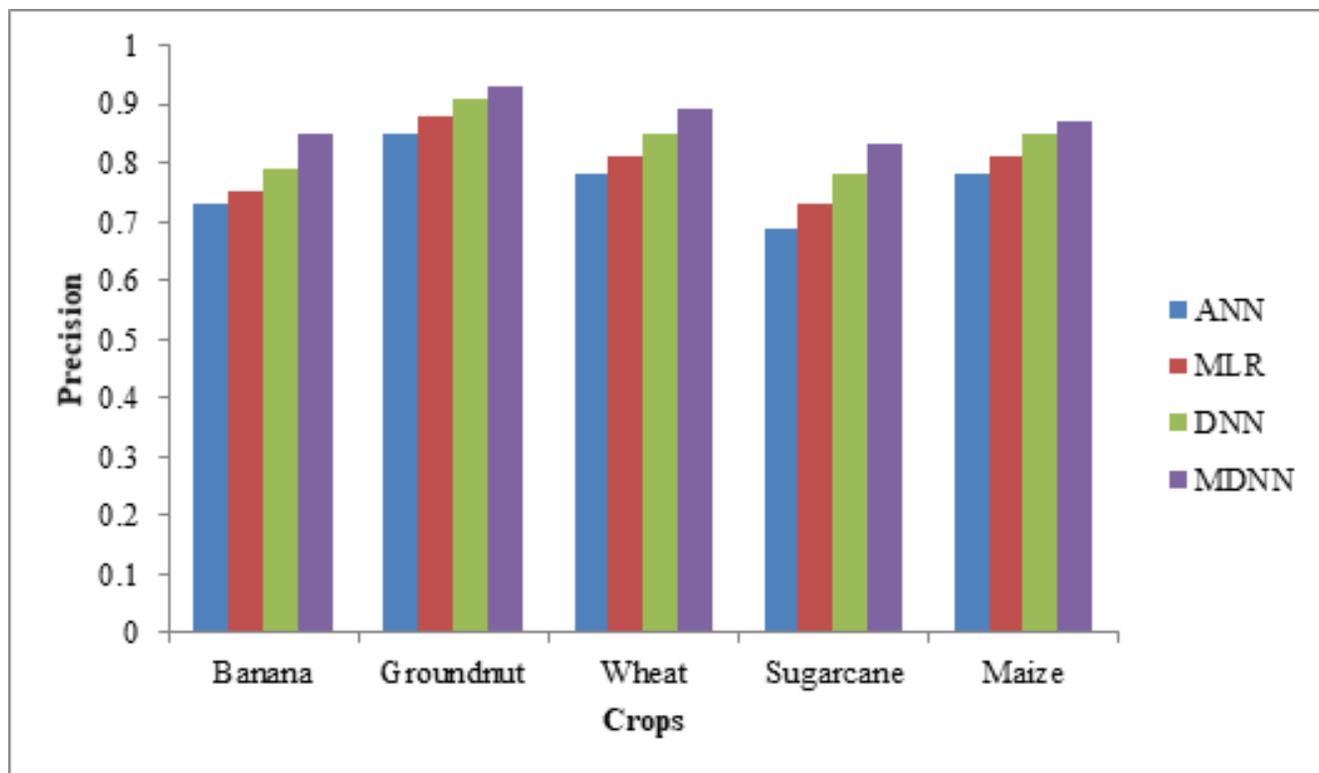


Fig 4. Comparison of precision for DNN and MDNN

Figure 4 shows the precision of DNN and MDNN for predicting 5 categories of crop yield. The precision of MDNN is 7.84%, 2.2%, 4.82%, 7.05% and 18.95% greater than DNN for banana, groundnut, wheat, sugarcane and maize crops respectively. The lReLU permits non-zero gradient during idle state of node which enhance the precision of MDNN for crop yield prediction. Thus, it notices the MDNN has a high precision than DNN for predicting all considered crop yields.

4.3 Recall

It is the amount of correctly predicting the crop yields over TP and FN.

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

Table 3 shows the recall outcomes obtained from DNN and MDNN for predicting 5 different crops yield.

Table 3. Comparison of DNN and MDNN in terms of recall

Methods	Banana	Groundnut	Wheat	Sugarcane	Maize
ANN	0.81	0.82	0.78	0.82	0.80
MLR	0.84	0.87	0.87	0.83	0.88
DNN	0.89	0.9	0.91	0.9	0.9
MDNN	0.913	0.926	0.931	0.927	0.928

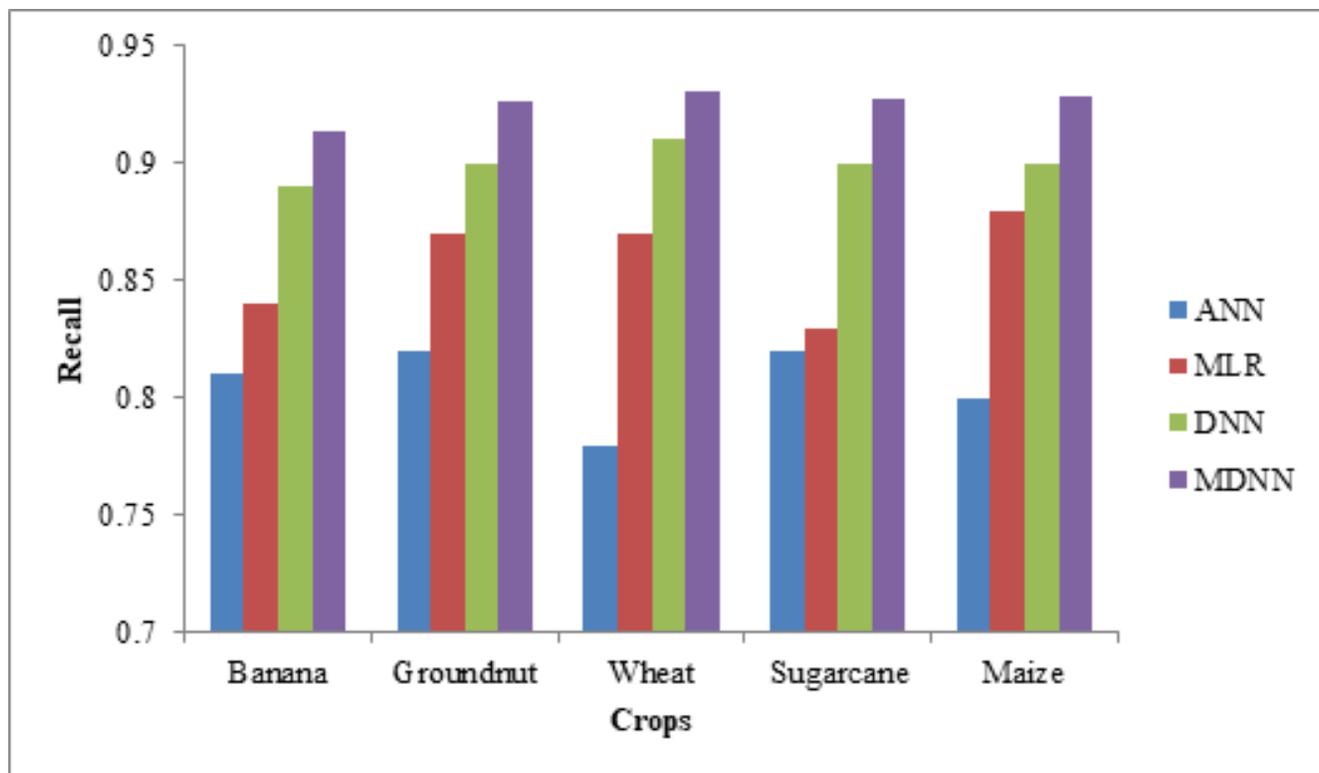


Fig 5. Comparison of recall for DNN and MDNN

Figure 5 shows the recall of DNN and MDNN methods for 5 different crops yield prediction. The recall of MDNN is 2.58%, 2.89%, 2.31%, 3% and 3.11% greater than DNN for banana, groundnut, wheat, sugarcane and maize crops respectively. The MDNN used the climate, weather, soil parameters and impact of climate change for crop yield prediction which enhance the recall rate of MDNN. So, it observes the MDNN has a high recall than DNN-based crop yield prediction for all variety of crops.

4.4 F-measure

It is the harmonic average of precision and recall.

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

Table 4 shows the f-measure outcomes obtained from DNN and MDNN for 5 variety of crops.

Table 4. Comparison of DNN and MDNN in terms of F-measure

Methods	Banana	Groundnut	Wheat	Sugarcane	Maize
ANN	0.71	0.78	0.80	0.76	0.74
MLR	0.76	0.84	0.83	0.82	0.78
DNN	0.78	0.91	0.88	0.86	0.80
MDNN	0.824	0.924	0.915	0.912	0.825

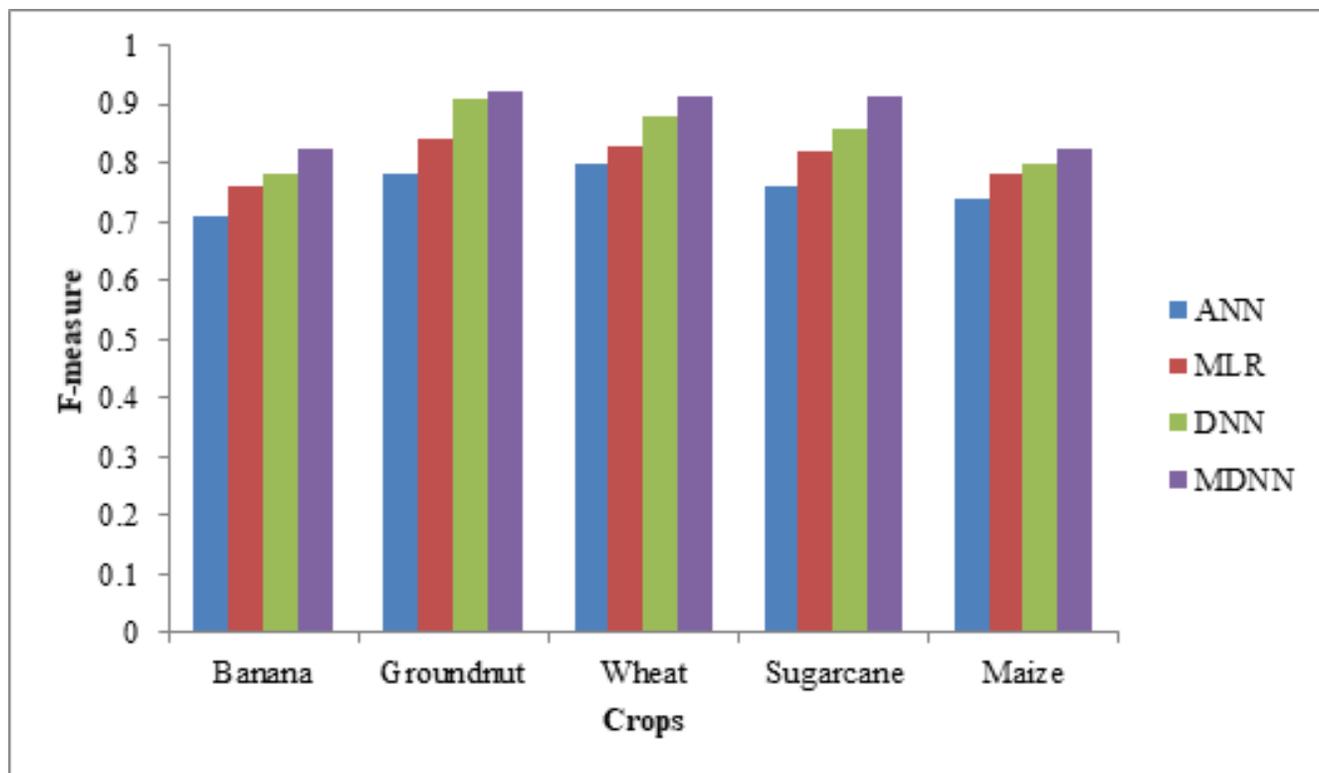


Fig 6. Comparison of F-measure for DNN and MDNN

Figure 6 shows the f-measure of DNN and MDNN for predicting the yield of 5 crops. The F-measure of MDNN is 3.11%, 1.54%, 3.98%, 6.05% and 3.13% greater than DNN for banana, groundnut, wheat, sugarcane and maize crops respectively. By improving the precision and recall rate, the F-measure of MDNN is also improved for crop yield prediction. Thus, it indicates the MDNN increases the F-measure compared to DNN for all considered crop yield prediction.

5 Conclusion

In this article, MDNN is proposed for accurate prediction of crop yield based on weather, climate, soil and impact of climate parameters. The effect of weather conditions on the crop yield is calculated based on the GDD measure using statistical inference. A neural network is used to predict the weather from the previous year of weather parameters. The predicted weather, soil, climate and impact of climate parameters are given as input DNN for predicting the crop yield. lReLU is used in the activation units of MDNN which enhance the accuracy of predicting the crop yield. The investigational outcomes proved that the MDNN achieves an enhanced performance compared to the DNN for predicting the crop yields. But, learning representations through a hierarchy of features increasing complexity, there are no guarantees about the quality of the last hidden representation. The learning process may fail or it could not provide an optimal representation especially for medium-sized datasets. So, the future extension of this research will handle the problem of guarantees about the quality of the last hidden representation.

References

- 1) Nosratabadi S, Karoly S, Beszedes B, Felde I, Ardabili S, Mosavi A. Comparative analysis of ANN-ICA and ANN-GWO for crop yield prediction. In: RIVF International Conference on Computing and Communication Technologies (RIVF);vol. 2020. 2020;p. 1–5. Available from: <https://doi.org/10.1109/RIVF48685.2020.9140786>.
- 2) Khaki S, Wang L, Archontoulis SV. A CNN-RNN Framework for Crop Yield Prediction. *Frontiers in Plant Science*. 2020;10. Available from: <https://dx.doi.org/10.3389/fpls.2019.01750>.
- 3) van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*. 2020;177. Available from: <https://dx.doi.org/10.1016/j.compag.2020.105709>.
- 4) Crane-Droesch A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*. 2018;13(11). Available from: <https://dx.doi.org/10.1088/1748-9326/aae159>.

- 5) Champaneri M, Chachpara D, Chandvidkar C, Rathod M. Crop yield prediction using machine learning. *International Journal of Science and Research*;2020(4):1–4.
- 6) Khaki S, Wang L. Crop Yield Prediction Using Deep Neural Networks. *Frontiers in Plant Science*. 2019;10:1–10. Available from: <https://dx.doi.org/10.3389/fpls.2019.00621>.
- 7) Gopal PSM, Bhargavi R. A novel approach for efficient crop yield prediction. *Computers and Electronics in Agriculture*. 2019;165. Available from: <https://dx.doi.org/10.1016/j.compag.2019.104968>.
- 8) Khaki S, Khalilzadeh Z, Wang L. Predicting yield performance of parents in plant breeding: A neural collaborative filtering approach. *PLOS ONE*. 2020;15(5). Available from: <https://dx.doi.org/10.1371/journal.pone.0233382>.
- 9) Wang X, Huang J, Feng Q, Yin D. Winter Wheat Yield Prediction at County Level and Uncertainty Analysis in Main Wheat-Producing Regions of China with Deep Learning Approaches. *Remote Sensing*. 2020;12(11). Available from: <https://doi.org/10.3390/rs12111744>.
- 10) Park JK, Das A, Park JH. Integrated model for predicting rice yield with climate change. *International Agrophysics*. 2018;32(2):203–215. Available from: <https://doi.org/10.1515/intag-2017-0010>.
- 11) Afrin S, Khan AT, Mahia M, Ahsan R, Mishal MR, Ahmed W, et al. Analysis of soil properties and climatic data to predict crop yields and cluster different agricultural regions of Bangladesh. In: and others, editor. IEEE/ACIS 17th International Conference on Computer and Information Science. 2018;p. 80–85. Available from: <https://doi.org/10.1109/ICIS.2018.8466397>.
- 12) Adisa O, Botai J, Adeola A, Hassen A, Botai C, Darkey D, et al. Application of Artificial Neural Network for Predicting Maize Production in South Africa. *Sustainability*. 2019;11(4). Available from: <https://dx.doi.org/10.3390/su11041145>.
- 13) Rao PR, Gowda SP, Prathibha RJ. Paddy Yield Predictor Using Temperature, Rainfall, Soil pH, and Nitrogen. In: Emerging Research in Electronics. Springer. 2019;p. 245–253. Available from: https://doi.org/10.1007/978-981-13-5802-9_23.
- 14) Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of the 30th International Conference on Machine Learning;vol. 30 of 1. 2013;p. 1–6.
- 15) Xu B, Wang N, Chen T, Li M. Empirical evaluation of rectified activations in convolutional network. 2015.