International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Novel Integration of Analytical and Simulation Approaches for Wheat Yield Prediction with Deep Learning Models

Nandini Babbar[1], Dr. Ashish Kumar[2], Dr. Vivek Kumar Verma[3]

Submitted: 27/01/2024 **Revised**: 05/03/2024 **Accepted**: 13/03/2024

Abstract: Food, water, and air are necessary for everyone's survival. Since the population is increasing continuously, in the same way, the agricultural industry needs to adopt various new techniques to fulfil the requirements. Changes in climate conditions have an adverse effect on the crops with the healthy existence of agriculture. Traditional farming methods are not fulfilling the needs. Since machine learning and various simulation models are used for early prediction of crop yield. Intelligent algorithms detect at early stages the production of crops and for the same, we can use the required material to increase the healthy production. Also, there are various simulation models which help in forecasting the yield based on the previous data. In this paper, there is a discussion about DSSAT and WOFOST simulation model for yield prediction. Also, here in this paper a new model is introduced using CNN and LSTM for better yield prediction named as Supervised Deep convolutional long-short term memory which gives R^2 =0.91, MAPE=1, MSE=,8 RMSE=2, MAE= 0.01.

Keywords: *Agriculture, Deep Learning, Decision support system for agrotechnology transfer, Convolutional Neural Network, long short-term memory.*

1. Introduction

India is well known for agricultural growth and development. It is extremely important for the entire economy. The major objective of agricultural production is to achieve maximum yield production at minimum expense. Agri-technology and precision farming, in the current era is also termed as digital agriculture, and this new data intensive discipline approach have a good and beneficial impact on the environment [1]. In India agriculture accounts for even greater than 70% of total land area. Extremely hard work is done by the farmers to increase the agricultural production with limiting effect on economic growth, and for that they are putting their efforts from early morning on daily basis with huge stress to contribute to the development of their country. Though, the farmers' current farming practices are behind the times and do not meet the needs for current requirements with desired outputs. They have adapted to their demands because the ancient inhabitants cultivated crops on their individual land. Consequently, a variety of animals and birds, including humans, have employed the naturally occur-ring commodities that are grown.

People in modern times don't understand how important it is to grow crops in the right place and at the right time. These cultivation practices are also changing the seasonal climate, endangering vital resources like air, water, and land, and leading to food insecurity. We have concluded that there is no suitable technology or solution to assist us in overcoming the current situation after assessing all of these difficulties and problems, including weather, temperature, and a range of other factors.

Meteorological parameters, such as humidity, temperature, and rainfall patterns, have a significant impact on agricultural yield growth. In this framework, farmers must seek assistance on forecasting weather conditions to optimize crop productivity. Agriculture's throughput has yet to reach its full potential due to a lack of techno-logical participation. Every farmer wants to know what kind of yield he may expect during the harvest season, so yield prediction is crucial for them. Weather variables, on the other hand, are major factors of crop production and can throw off conventional yield predictions by altering some of the soil and ecological variables that affect crop development.

From the agricultural area, there are several crops, but wheat and rice are the most necessary crops. India is the world's second-largest wheat producer, with 29.7 mil-lion hectares under agricultural production. Haryana, Punjab, Uttar Pradesh, and Madhya Pradesh are the states which produce the highest wheat.

[1] *Manipal University Jaipur*
ORCID ID : 0009-0008-6437-5326
[2] *Manipal University Jaipur*
ORCID ID : 0000-0002-6644-9754
[3] *Manipal University Jaipur*
ORCID ID : 0000-0003-1998-8736
** Corresponding Author Email: nandinibabbar1@gmail.com*

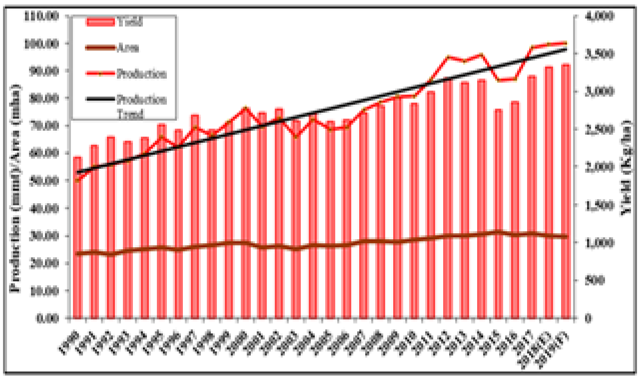

Fig. 1. To show the Graph showing the Winter Wheat Yield Forecast for 2019 (Marketing Year 2019–20) from FAS/New Delhi, the Indian government's Ministry of Agriculture and Farmers Welfare.

2. Literature Review

Bhatnagar, R., et al. [2] has proposed a simulation model with machine learning and data mining methods used to predict rice yield predictions. Predictability of crop yields is considered to depend on the weather and soil conditions. The results of this study reveal that projected data from a random forest and Crop Yield Estimates have (R^2 = 0.67) in addition to RMSE of 281 kg/ha. Following the simulation, the authors employed machine learning methods to predict the yield of crops depending on known variables as well as the responses of previous calculations. Based on the pro-duction tree, the decision tree is used to form a tree structure and estimate yield.

Gohain, et al.[3], The Provincial Crop Production Assessment Program that has been investigated, employs a crop simulation approach and a framework intended to enhance the effective application of the DSSAT simulations model at the local, state, and national levels. This approach was used to analyse monitored files at the neighbourhood, region, ranging and territorial levels by utilizing the different properties of the soil, commodity growing and flowering records, and cultivating files.

X.E. Pantazi, [4] suggested using three supervised learning algorithms—the XY-fusion network (XYF), the Supervised Kohonen Network (SKN), and the Counter-propagation Artificial Neural Network (CPANN)—to associate accurate agricultural data with the frequency categories of crop production. The algorithms make use of pre-existing data collected through various kinds of vegetation and soil measurements using unsupervised neural networks algorithm design. The production region comprises anticipated iso-frequency yield categories from three training models, and the anticipated input nodes include different variables impacting soil properties and the satellite normalized differentiation vegetation index (NDVI). The study concluded that, considering a complex relationship between boundaries and yield, the yield forecast depends on a cross-sectional evaluation of the SKN model of the lower harvested subcategory above 91%. This is an exceptionally high level of accuracy.

Kadir, et al. [5] has attempted to determine which model—the CERES-Wheat model, the SIRIUS method, and the AFRCCHEAT2 model—is the best at forecasting the production of wheat using a variety of data sources. In this study, an artificial neural network (ANN) based on backpropagation using a layer structure known as Multi-Layer Perceptron (MLP) was used to construct a wheat yield prediction model. Input parameters include sun, snow, rain, and temperature, and the data spans from 1997 to 2007. The model output parameter makes use of wheat yield data spanning the years 1997 through 2007. Instruction, verification, and assessment are the three distinct groups into which the data is divided. This MLP could predict wheat yield with 98 percent accuracy.

Cadenas, et al. [6] has done their work on wheat crop development and sustainability in main wheat-growing regions, and India is required to present a whole image of wheat production behaviour in the future. A hybrid ARIMA – ANN model is utilized for predicting the speed of wind in a separate location of Mexico. Because every time series cannot be considered totally linear or indirect due to its complexity, it is required to appropriately read both straight and indirect sections. To capture the whole behaviour of various sorts of data series, hybrid prediction algorithms are used. Therefore, there is great potential for our MLP-based wheat harvesting model as a tool that can generate precise estimates in comparison to wheat yields and can also be used to other crops.

Cao, et al. [7] has compared the predictions for winter wheat yield in the China region using three deep learning algorithms and machine learning. While deep learning algorithms include long short-term memory networks, convolutional neural net-works, and deep neural networks, machine learning methods include Random Forest. Previously, for the purpose of predicting yield, they integrated the forecast wheat yields, with more and more people using machine learning as well as algorithms based on deep learning. These models make use of huge datasets, such as images from satellites, variations in the weather, soil attributes, and archival production data. A variety of data sources should be used to ensure reliable projections, according to the literature. Among these sources are historical yield records, soil data, meteorological data, and satellite-based remote sensing information. Although the results are encouraging, there are still difficulties in precisely forecasting wheat yields. The constantly changing characteristics of systems related to agriculture, the complexity of models, and data dependability are some of the constant obstacles. The effectiveness of models may be hampered by a lack of ground-truth information needed for both validation and training, which calls for the creation of imaginative techniques for data collecting and evaluation.

Feng, et al. [8] has compared and investigated the growing trend of using neural networks to combine spatial and temporal factors for more precise and targeted pre-dictions. They have emphasized the importance of adding temporal and spatial elements to models for prediction in agricultural research. Accurate yield predictions require complex methodology because of the temporal variations in soil, climate, and management techniques that occur in different places. Large datasets with com-plicated patterns can be effectively captured by neural networks, which have shown to be useful tools. They are appropriate for agricultural yield prediction because of their capacity to learn from past data, such as meteorological data, soil properties, and satellite imagery. Due to its ability to take spatial variability into account, geo-graphically weighted models have drawn more attention. These models can improve forecasts by varying the weights given to nearby places according to local variables. These models improve forecasts in areas with varying agricultural practices and environmental conditions by allocating varying weights to nearby sites based on local factors. A more complex framework for predicting winter wheat output is provided by the innovative method of fusing neural networks with temporal and spatial weighting algorithms.

Korohou, et al. [9] has put out a study project that examines wheat yield estimates or evaluations based on an amalgamation of morphological properties, comprising several wheat constituents. By comparing the perimeter of the harvest of wheat with the total length of the stalk, the R2 values of 0.9609 and 0.9779, correspondingly, were used to evaluate the use of the three morphological characteristics (dimension, broadness, and perimeter) that were included utilizing a computerized detection approach. Six models of regression were created based on the features that were recovered. With R^2 of 0.9893 and RMSE of 0.0684 mm in evaluating dry grain out-put with wet wheat weight as predictions, the linear regression is dependent on the moisture content of the stemmed along with every other mathematical model that were examined.

3. Yield Prediction Models

Predicting crop yield is a foremost and an important task for the decision makers at regional as well as at national levels for correct and rapid decision making for both economically and to fulfil everyone’s daily needs. We need to work with an intelligent system that can predict the yield based on all the parameters that are affecting it. Accurate yield forecasting models assist farmers in determining what to grow and when. Numerous methodologies are there for yield forecasting. This article explains what the various models are used for yield prediction and the machine learning algorithms used.

3.1. Statistical Model

Since the population is growing at an exponential rate, crop productivity is decreasing due to unforeseen rain, windstorms, or predictable rain daily. Our precision agriculture system collects present statistics based on climate, quality of air, characteristics of soil, crop development, and crop infection, as opposed to traditional agriculture, which entails establishing or gathering the crop according to a scheduled plan. These predictive analytics can be applied in the agricultural area to make better decisions. Precision agriculture is the process of evaluating previous year yield values, comprehending current environmental requirements, and analyzing data that computes variances in soil in addition to crops surrounded by agricultural areas. Predictive farming is focused on determining whatever has occurred, assessing what would be occurring now, and regulating the situation to meet future agricultural needs.

3.2. Simulation Model

A Crop Simulation Model (CSM) determines growth of crop as well as its development procedure based on process analysis and performance evaluations in naturalistic environments [14]. Figure 2 indicates the various parameters used in simulation model and statistical model and shows the complete work process. They are dynamic models that seek to simulate crop growth and development by utilizing the underlying mechanism of crop and soil processes. Simulation methods help in investigating production differences in a variety of crops, without the requirement for costly and time-consuming field research. Crop simulation models are extremely valuable because they serve as a link between crop process analysis and performance evaluation.

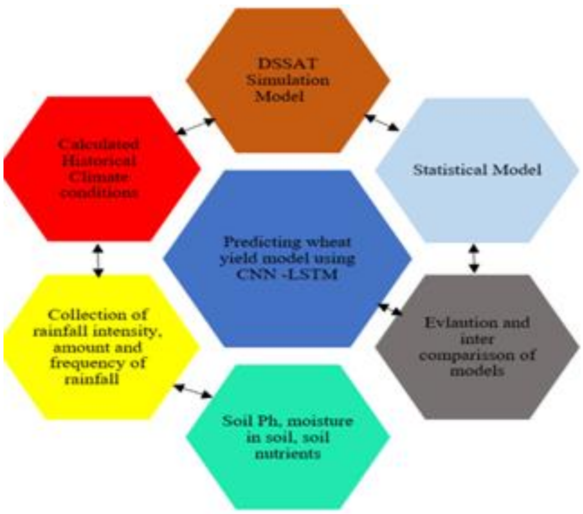

Fig. 2. Common parameters used in Simulation, Statistical model

3.3. Decision Support System for Agrotechnology Transfer (DSSAT Simulation Model)

It is basically used to determine the simulation of crop development and production over time, and soil water, carbon, and nitrogen processes and management strategies. For soil-plant-atmosphere dynamics, DSSAT is supported by several programs and applications. Academics, educationalists, counsellors, cultivators, isolated traders, policies as well as decision makers, and many more have employed DSSAT in more than 174 countries for 30 plus years. Various dataset services and applications for different models of crop includes weather conditions, characteristics of soil, genetic factors and all that is supported by DSSAT. The yield in DSSAT model is a function of crop management and observational experiment data.

We are having metrological and soil parameters in consideration. Both the data were in csv file. In Figure 3 that csv data is converted into the grid format.

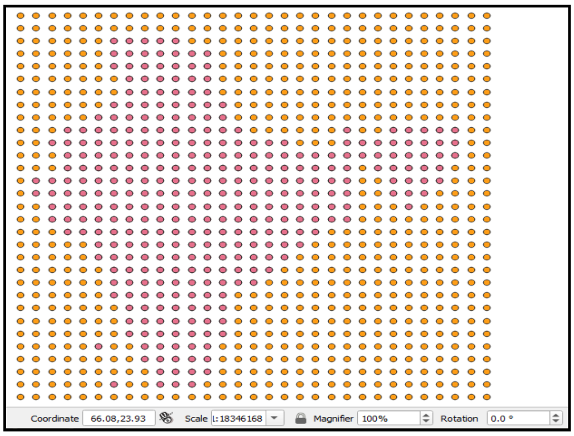

Fig. 3. Conversion .csv data in grid format for analysing in DSSAT simulation model

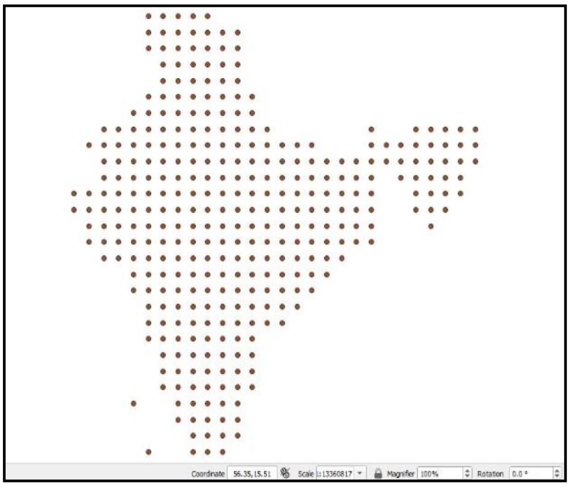

Fig. 4. Graphical representation for longitude and latitude for csv data

Initially before working on DSSAT model, we get data from all over the world in the csv format. Then we must convert the data in particular grid format and after that we used to put longitude and latitude of a particular region of India then we will get the below output. After conversion of data in grid format the graphical structure can be seen for specific longitude and latitude as shown in figure 4.

4. Proposed Model

The proposed model Supervised deep convolutional long-short term memory (SDC-LSTM) integrated the concepts from deep learning models. Various algorithms are used for prediction of yields. In this work, concepts of Convolutional neural network and long short-term memory algorithms with enhanced features are used for better prediction of wheat yield.

This model proposes effective deep learning-based methods to predict wheat yield, namely, the Superior Deep Convolutional Long-Term Short-Term Memory (SDC-LSTM) Neural Network Framework. Weather data, previous year's wheat yield, and soil properties data are collected in the first level.

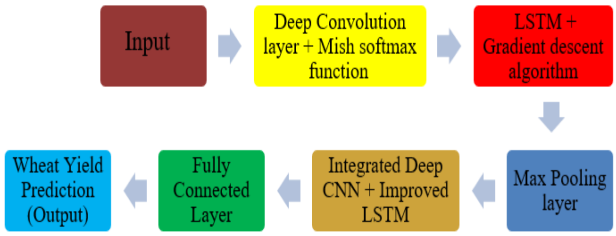

Fig. 5. To show the pre-processed data are then fed into the SDC-LSTM framework

The pre-processed data are then fed into the SDC-LSTM framework, which has different features such as dynamic and static data. Furthermore, Superior Deep Convolutional Neural Network is introduced for detecting dynamic features, in which our research proposes the Mish SoftMax function in the convolution layer for detecting weather data, wheat yield data of previous years to improve accuracy, max pooling is introduced before the flatten layer, which gives the network superior stability. In addition, Improved Long Term Short Memory is proposed to detect static features (soil data). The gradient descent algorithm is used early to stop overfitting in the network, improving the network with training data. Furthermore, dynamic and static features are combined. The fully associated level network receives the integrated data after which it normalizes and assesses it. Finally, the wheat yield is predicted by our SDC-LSTM framework.

5. Result and Discussion

In this model, Rajasthan's data is considered as referenced data. A total of 50 epochs are used for the implementation process. 80:20 ratio is considered for the training and testing purpose. Figure 6 indicated that as the number of epochs increased the mean absolute error is reduced for the training as well as testing model and in this model, the training proposed method is very less when compared to the testing proposed model. Figure 7 shows how much loss we get by implementing the purposed model.

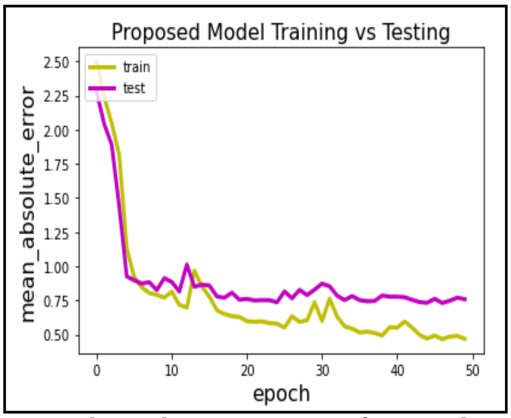

Fig. 6. To show the SDC-LSTM framework Model Training vs Testing

By executing the designed model for wheat yield prediction considering the data for Rajasthan state, the result with minimum value for Mean Absolute Error i.e. almost equal to 0.0001 as shown in Figure 8. As we used various libraries of python as well in our work, we are able to calculate the other parameters for describing the better work and the value received for the parameters by using Supervised Deep Convolutional Long-Short Term Memory model are R^2 i.e. statistical measure of goodness of fit is 0.91, MAPE i.e. Mean absolute percent-age error to measure the accuracy is 1 , MSE i.e. Mean square error to find the loss in function is 8, RMSE i.e. Root mean square error to find the difference between actual and predicted values of model is 2 , MAE i.e. Mean absolute error to find the magnitude of difference between the actual and predicted value is almost negligible.

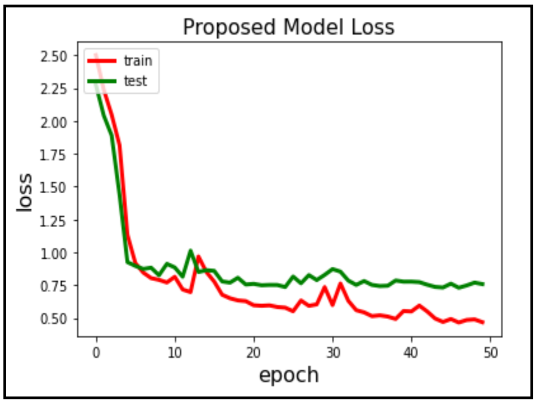

Fig.7. To show the SDC-LSTM framework Model Loss

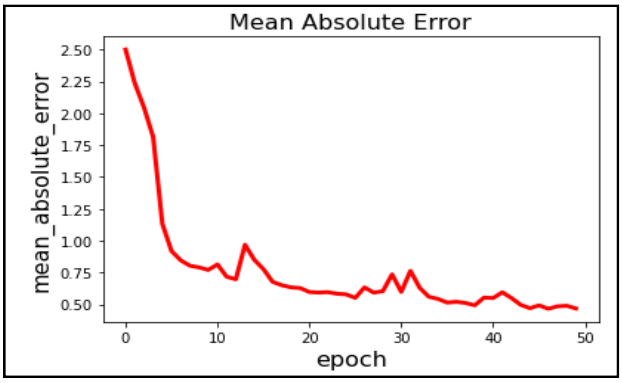

Fig.8. To show the MAE of SDC-LSTM framework Model

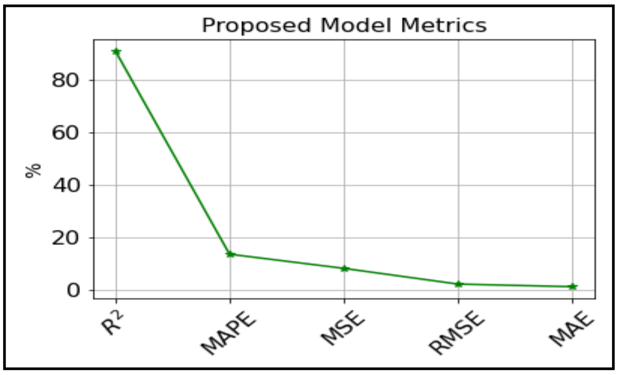

Fig.9. SDC-LSTM framework Metrics

6. Conclusion

It is concluded that various supervised, unsupervised machine learning algorithms, deep learning algorithms worked to detect the yield prediction of yield. Simulation models like Decision Support System for Agrometry Technology and World Food Studies model also predicted the yield based on certain parameters. There are various parameters, for example weather conditions, soil characteristics and crop management that affect the production and too the yield of any crop. Till now the work done for wheat yield prediction was done by considering individual factors or combination of two. In the future, multi parameters can be considered to increase the efficiency for the yield prediction. By gathering and pre-processing soil and weather data, the designed method forecasts wheat production. A framework with better deep convolutional long short-term memory receives the pre-processed data. We are using an improved long short-term strategy and gradient descent technique to get better accuracy by removing overfitting and early stopping issues in the network. The completely linked layer receives data from the integrated layer to normalize. Finally, the research successfully forecasts wheat yield with efficient results.

Author contributions

Nandini Babbar: Conceptualization, Methodology, Simulation modelling, Field study, Draft preparation **Dr Ashish Kumar:** Data curation, Editing draft preparation, Statistical modelling, Validation, Dr **Vivek Kumar Verma:** Visualization, Validation, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Liakos, Konstantinos G., Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. "Machine learning in agriculture: A review." Sensors 18, no. 8 (2018): 2674.

[2] Bhatnagar, Roheet, and Ganesh Borpatra Gohain. "Crop yield estimation using decision trees and random forest machine learning algorithms on data from terra (EOS AM-1) & Aqua (EOS PM-1) satellite data." Machine Learning and Data Mining in Aerospace Technology (2020): 107-124.

[3] Gohain, G. B., K. K. Singh, R. S. Singh, and Priyanka Singh. "Regional Crop Yield Estimation System (RCYES) using a crop simulation model DSSAT V4. 7: concept, methods, development, and validation." (2020): 33-38.

[4] Pantazi, Xanthoula Eirini, Dimitrios Moshou, Thomas Alexandridis, Rebecca Louise Whetton, and Abdul Mounem Mouazen. "Wheat yield prediction using machine learning and advanced sensing techniques." Computers and electronics in agriculture 121 (2016): 57-65.

[5] Kadir, Muhd Khairulzaman Abdul, Mohd Zaki Ayob, and Nadaraj Miniappan. "Wheat yield prediction: Artificial neural network-based approach." In 2014 4th International Conference on Engineering Technology and Technopreneuship (ICE2T), pp. 161-165. IEEE, 2014.

[6] Cadenas, Erasmo, and Wilfrido Rivera. "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model." Renewable Energy 35, no. 12 (2010): 2732-2738.

[7] Cao, Juan, Zhao Zhang, Yuchuan Luo, Liangliang Zhang, Jing Zhang, Ziyue Li, and Fulu Tao. "Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine." European Journal of Agronomy 123 (2021): 126204.

[8] Feng, Luwei, Yumiao Wang, Zhou Zhang, and Qingyun Du. "Geographically and temporally weighted neural network for winter wheat yield prediction." Remote Sensing of Environment 262 (2021): 112514.

[9] Korohou, Tchalla, Cedric Okinda, Haikang Li, Yifei Cao, Innocent Nyalala, Lianfei Huo, Mouloumdèma Potcho, Xiang Li, and Qishuo Ding. "Wheat grain yield estimation based on image morphological properties and wheat biomass." Journal of Sensors 2020 (2020): 1-11.

[10] Ray, Deepak K., Navin Ramankutty, Nathaniel D. Mueller, Paul C. West, and Jonathan A. Foley. "Recent patterns of crop yield growth and stagnation." Nature communications 3, no. 1 (2012): 1293.

[11] Le Mouël, Chantal, Elodie Marajo-Petitzon, Patrice Dumas, Olivier Mora, Stéphane Manceron, and Agneta Forslund. "Agrimonde-Terra foresight: Land use and food security in 2050: Scenarios' simulation results." (2016): 28-p.

[12] Michel, Lucie, and David Makowski. "Comparison of statistical models for analyzing wheat yield time series." PLoS One 8, no. 10 (2013): e78615.

[13] Kumar, Ajay, P. K. Deepankar, M. PK, and A. Kumar. "Wheat yield forecasting in Haryana: A time series approach." Bull. Env. Pharmacol. Life Sci 8, no. 3 (2019): 63-69.

[14] Yadav, S., A. Misra, S. Mishra, and Vyas Pandey. "Crop growth simulation models (InfoCrop v. 2.1, DSSATv4. 5, WOFOSTv1. 5 and Cropsytv 4.19) software." Water Energy Secur. Arena Clim. Chang 34 (2018): 456-469.

[15] Feng, Quanlong, Jiantao Liu, and Jianhua Gong. "Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier—A case of Yuyao, China." Water 7, no. 4 (2015): 1437-1455.

[16] Ali, Iftikhar, Felix Greifeneder, Jelena Stamenkovic, Maxim Neumann, and Claudia Notarnicola. "Review

of machine learning approaches for biomass and soil moisture retrievals from remote sensing data." Remote Sensing 7, no. 12 (2015): 16398-16421.

[17] Ip, Ryan HL, Li-Minn Ang, Kah Phooi Seng, J. C. Broster, and J. E. Pratley. "Big data and machine learning for crop protection." Computers and electronics in agriculture 151 (2018): 376-383.

[18] Jeong, Jig Han, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher, Kyungdahm Yun, Ethan E. Butler, Dennis J. Timlin et al. "Random forests for global and regional crop yield predictions." PloS one 11, no. 6 (2016): e0156571.

[19] Han, Jichong, Zhao Zhang, Juan Cao, Yuchuan Luo, Liangliang Zhang, Ziyue Li, and Jing Zhang. "Prediction of winter wheat yield based on multi-source data and machine learning in China." Remote Sensing 12, no. 2 (2020): 236.