iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.43191

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Crop Yield Prediction Using Machine Learning

Gagan M[1], Ganta Narendra Reddy[2], Kethan S. A[3], Kiran. K[4]
[1, 2, 3, 4]*School of Computer Science and Engineering, REVA University, Bangalore, Karnataka, India*

Abstract: ***Agriculture yield increase and agroindustry goods account for the majority of India's economy. They are the economic backbones of agricultural countries. Yield prediction is a crucial topic in agriculture. Any farmer wants to know what kind of harvest he may expect. Analyze the numerous connected parameters that are utilized to determine the alkalinity of the soil, such as location and pH value. Additionally, third-party apps such as APIs for weather and temperature, type of soil, nutrient value of the soil in that place, quantity of rainfall in that region, and soil composition are used to compute percentages of nutrients such as nitrogen (N), phosphorus (P), and potassium (K). To develop a model, each of these data properties will be evaluated and trained using various machine learning methods. The system includes a model that is precise and reliable in forecasting crop output and providing proper fertilizer ratio recommendations based on atmospheric and soil data. Farmers may also learn which crops are in great demand so that they can be readily produced. This application might be beneficial to all farmers who want to know which crops can be grown in various places or soils in order to optimize their profitability.***
Keywords: Crop Yield Prediction, Machine Learning, and Genetic Algorithm

I. INTRODUCTION

Food security has become more scientifically important as the world's population continues to expand. Distribution patterns and yield fluctuations of staple crops must be examined in order to preserve grain security, since they can enhance farm management and farm economic planning. Crop production is mostly estimated at the regional scale using vegetation indices collected from remote sensing data. The strategies may be classified into single index and multiple index strategies. The first technique employed an empirical model to determine crop production using vegetation indices such as the normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), and normalized difference water index (NDWI) during important phenology ages in Cloud Computing [12]. The fundamental objective of the second strategy was profitability and long-term viability. Precision agriculture's primary purpose is to increase crop quantity. Precision agriculture entails employing information technologies to better understand the crop. The other technique employed time series vegetation indices to enhance a crop growth process model, which was often used to estimate agricultural production in advance. The approach using several indices was found to be slower than the method using only one index. This is an algorithm that seeks to develop the most efficient model to anticipate the crop's yield, thus test out multiple methods and compare them to see which one has the least error and loss, then choose that model. Nonetheless, the goal is to create a model that can reliably predict agricultural productivity while also allowing farmers to understand which crops are in great demand. so that they may be easily cultivated This service may be useful to all farmers who wish to know which harvests may be grown in particular places or soils in order to attract potential clients. Fig. 1 shows the Generic Architecture Diagram.

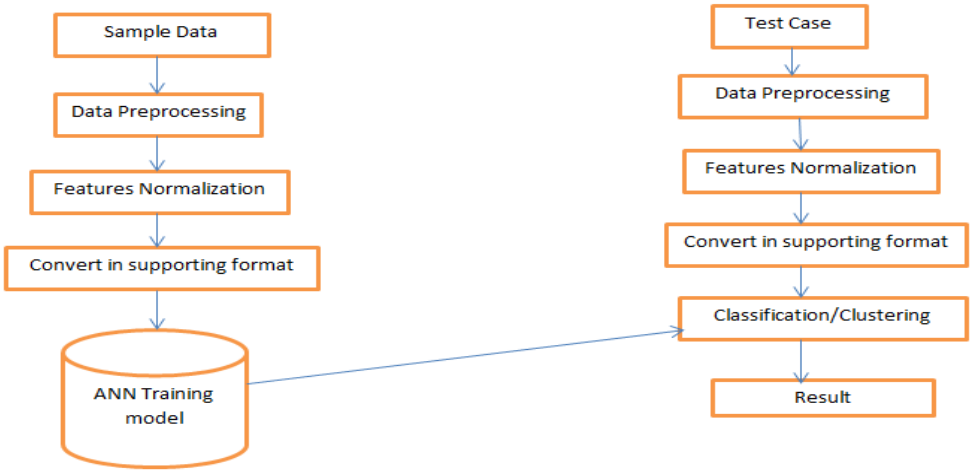


Fig. 1: Generic Architecture Diagram.

II. RELATED WORK

For Action Recognition, a variety of methodologies and methods have been developed. This section summarizes the previous efforts:

[Food pricing, food security, and land use decisions are all influenced by changes in global crop output [1]. Increased temperatures have a demonstrable detrimental effect on worldwide yields of wheat, maize, and barley. The estimate that warming since 1981 has resulted in yearly aggregate losses of these three crops is based on these sensitivities and observable climatic trends. The findings show that climate change is already having a detrimental influence on food production on a worldwide scale.

MODIS WDRVI built a crop yield estimation model utilizing MODIS data in a research report [2]. From 2000 to 2011, the model found geographical trends in maize final grain output across the US Corn Belt. Corn yields at the state level were reliably assessed, with a coefficient of variation of less than 10%. Around the East Coast, North Dakota, Minnesota, Wisconsin, and Missouri, on the other hand, the model tended to overestimate maize grain yield by scheduling the resources using VM Scheduling in Cloud Computing [13].

Establishing timely and highly accurate models for crop yield estimation is critical for crop management and decision-making, according to study article [3]. The goal of this work was to develop a technique for estimating spring maize production based on a crop growth model and the entropy method. The experiment was carried out in Northeast China's Jinchuan Farm. The entropy method (EM) was used to generate the combined weights of the single-temporal estimate models, and a combination forecasting (CF) model was created.

In research paper [4] Crop yield prediction is an essential task for the decision-makers at national and regional levels for rapid decision-making. Ground truth crop yields are collected from official statistics or directly from farmers. In this research paper the author trains the Model with Convolutional Neural Network and Recurrent Neural Network. Which capture the time dependencies of environmental factor CNN and RNN algorithm will determine pattern, correlate and discover knowledge and learn from Datasets. The output will be based on the past experience as it is artificial based Model [14].

In research paper [5] Climatic change trend might further negatively affect winter wheat production in the future. Winter wheat crop response to increase in soil fertility and chemical fertilizer was small during the simulation period. Temperatures, sunshine hours and relative humidity were all negatively related to grain yield of winter wheat. The climatic change trends in this area showed that the DTR and sunshine hours were declining. This type of climatic change trend could further negatively affect winter wheat production.

III. PROPOSED METHODOLOGY

Machine learning is used in the proposed system to train the provided dataset using the ANN algorithm. The dataset was obtained from the Kaggle website and is of Indian origin. The dataset must be pre-processed and the model trained. For GUI development, the Flask framework will be utilized. The yield with meteorological information and graphical representation will be provided in the GUI based on the specific year, state, and crop. Some future projections are also made to determine which crop is best for a specific state and year. The yield prediction is built on the three inputs: year, harvested area, applied area, average quantity applied, particular fertilizers applied, such as urea, and yield. In order to provide the necessary result, the data mining approach of multiple linear regression will be applied.

Every paper makes use of climatic variable conditions such as rainfall and sunlight, as well as agricultural factors such as soil type and fertilizer's (Nitrogen, Potassium, etc.) However, the problem is that we need to collect the data, have a third party make the prediction, and then explain it to the farmer, which takes a lot of time and effort on the farmer's part because he doesn't comprehend the science underlying these aspects.

To make things simple for the farmer, this paper utilizes simple parameters like the farmer's state and district, the crop he grows, and the season he grows it in.

1) *Normalized Difference Vegetation Index:* The normalized difference vegetation index (NDVI) is a simple graphical indication that may be used to determine whether or not the object being seen has living green vegetation using remote sensing readings, generally from a space platform. The estimation of crop yield on regional scale is mainly based on vegetation indices derived from remote sensing data. The methods could be divided into single index based and multiple indices-based strategies.
2) *ANN- Artificial Neural Network:* It's made to mimic the way the human brain analyses and processes data. The training results are used to predict the output values for given input parameters. Artificial Neural Networks (ANN) are algorithms that are based on brain processing and may be used to model complicated patterns and forecast issues.

Artificial neural networks (ANNs) make use of learning algorithms that may make modifications - or learn - on their own when fresh information is received. Fig. 2 shows the difference between Brain Neuron and ANN.

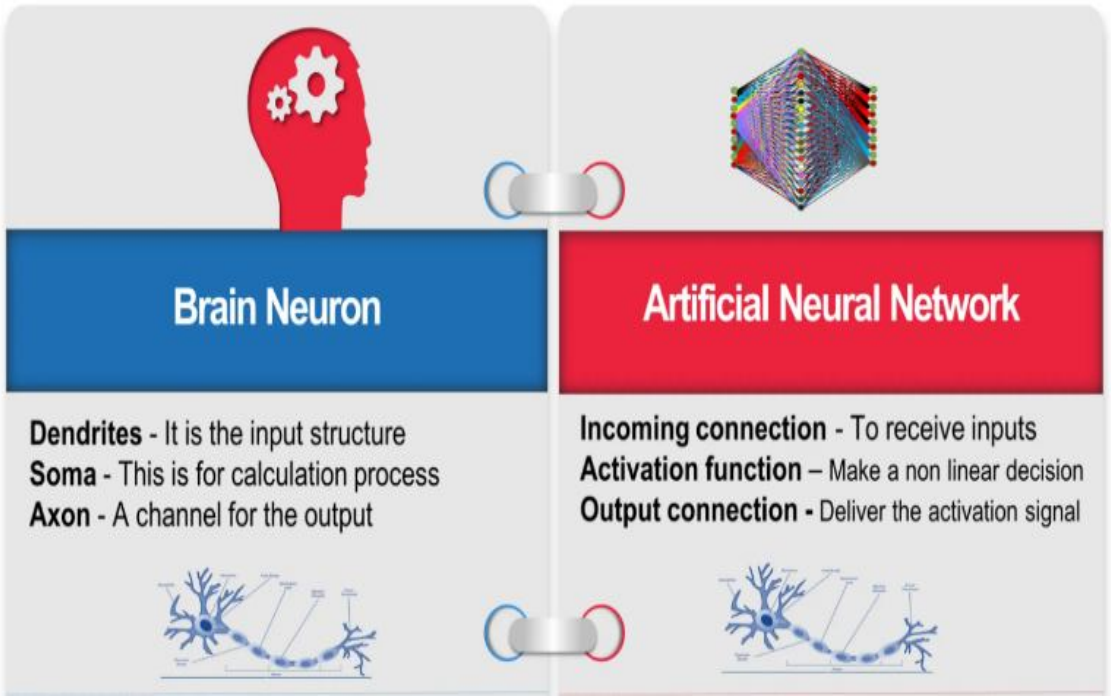


Fig. 2 shows the difference between Brain Neuron and ANN

- Gathering sample data
- Converting raw data into a suitable format
- Normalizing the range of independent variables or data aspects
- Converting raw data to a predetermined format for data transmission.
- ANN Technique was used to train the model.

The most critical step in tackling any supervised machine learning problem is gathering data. Your text classifier is only as good as the dataset it is trained on. To complete the training process as quickly as possible, the collected data should be analyzed and trained using the ANN algorithm. This has three levels.

3) *Convolution Layer:* This is where feature extraction occurs, where only relevant characteristics required by the computer are collected and undesired features are discarded, allowing the training phase to be completed quickly.
4) *Pooling Layer:* This reduces the amount of the data or image and produces a compressed document with critical elements required by the machine.
5) *Fully Connected Layer:* Here, the above data which we get from the previous layer will be fed to fully connected layer in a vector form. Then these compressed features will be split and get trained using ANN and will produce us the final output. In most cases, filter techniques are utilized as a preprocessing phase. The feature selection is independent of any machine learning methods. Instead, characteristics are chosen based on their performance in several statistical tests for connection with the result variable.

Data filtering is the process of choosing a smaller part of the data you have collected to show or analyze. Most of the time, but not always, filtering is temporary. The whole set of data is kept, but only a part of it is used in the calculation.
We begin by retrieving the dataset and performing pre-processing, which cleans the data and removes any features that just aren't useful for training the system. Only the critical features required by the system will be retrieved, and they will be converted into a format that only the system can recognize.

We use the ANN algorithm to train the data once we extract the specific attributes from the data. ANN will be performed using MS Excel to analyze the data acquired, using Yield as the dependent variable and the area harvested, area applied, average quantity applied, and different fertilizers applied as independent factors. Once the training is finished, we will classify the data by state, year, and crop. Finally, the input data will be recognized by the ANN algorithm, which will give us with the desired outcome. Graphs will also be provided dependent on the outcome. Fig 3. Shows the complete process of the proposed work and Fig 4. Shows the UML process of the proposed work.

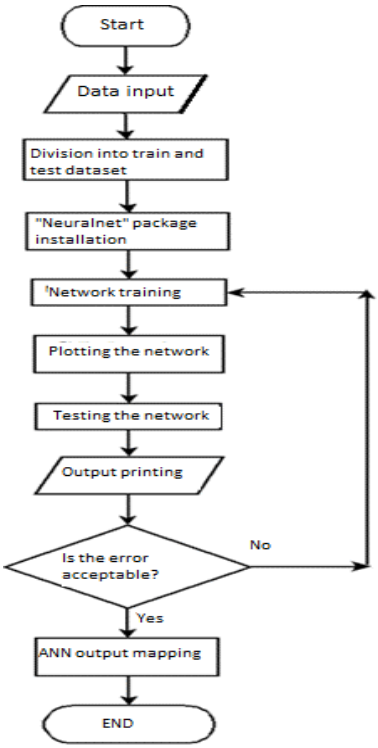


Fig 3. Shows the complete process of the proposed work.

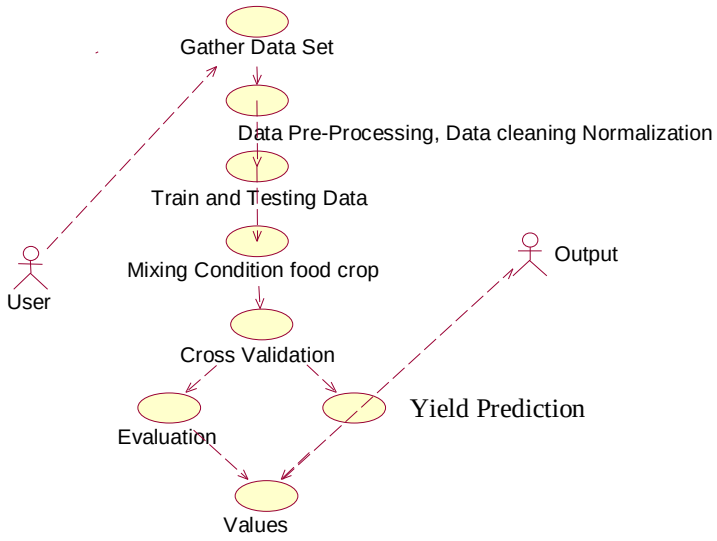


Fig 4. Shows the UML process of the proposed work

IV. RESULTS AND DISCUSSIONS

There are several research papers on crop yield prediction. But the objective is to create an efficient model for predicting the crop production. The other methods, which utilized multiple indices-based methods, was not time-efficient. An empirical model was used to determine crop production using vegetation indices such as the normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), and normalized difference water index (NDWI) during important phenology ages. Another method's primary objective was profitability and long-term viability. Precision agriculture's primary purpose is to increase agricultural yield. One technique employed time series vegetation indices to develop a crop growth process model, which was often used to estimate agricultural production in advance. The approach using several indices was demonstrated to be slower than the one using a single index. Fig 4. Shows the Home Page, Fig 5. Shows Crop in particular state in a particular year, Fig. 6 Shows the Graphical representation.

In our research the yield projection is based on the year, harvested area, applied area, average quantity applied, specific fertilizers used such as urea, and others, as well as the yield. In order to get the desired outcome, The crop output will be projected in kilograms per hectare based on the kind of crop, year, and state. The weather data may be shown in several ways, such as high temperature, low temperature, cloud cover, and so on. The visual representations will also be displayed for easier understanding of the output.



Fig 5. Shows the Home Page

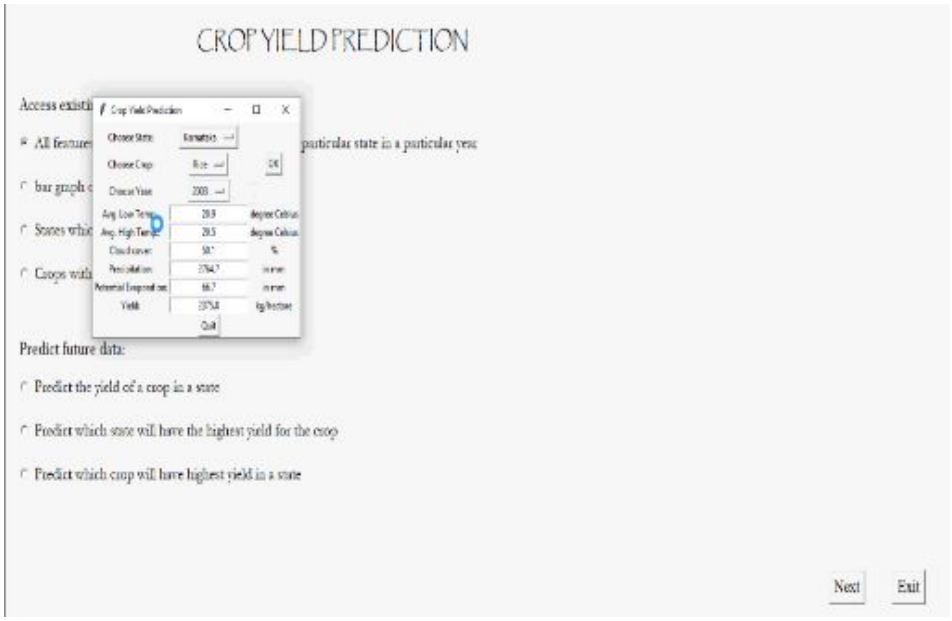


Fig 6. Shows Crop in particular state in a particular year

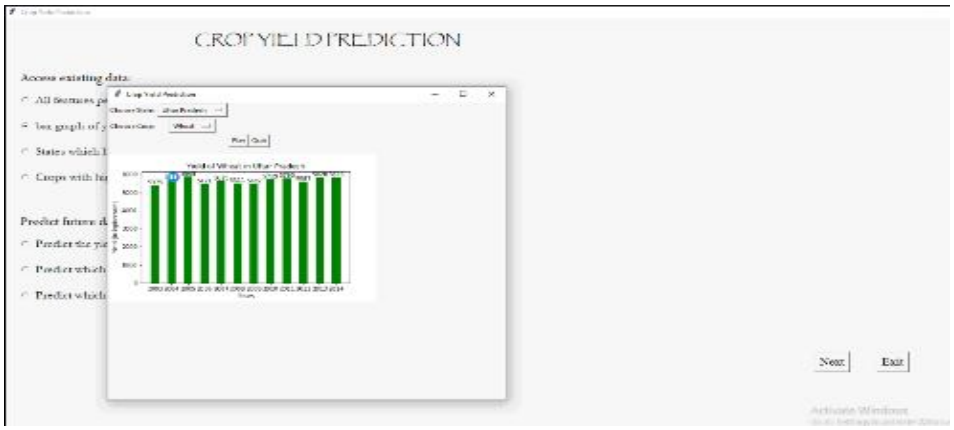


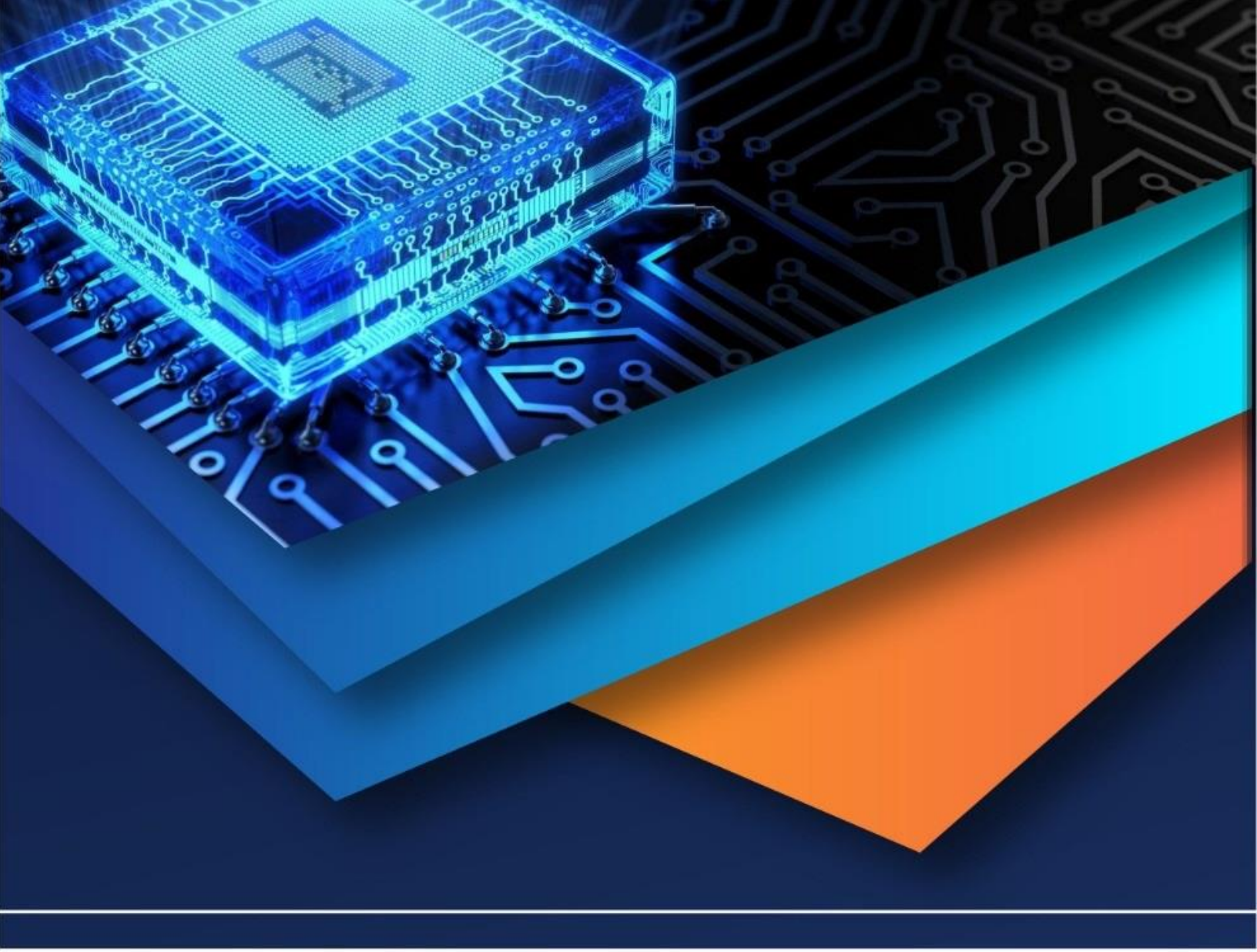
Fig. 7 Shows the Graphical representation

V. CONCLUSION

An observational model was used to determine crop production using vegetation indices during important phenology ages. The primary aim in this scenario was profitability and sustainability. Precision agriculture's primary purpose is to increase agricultural yield. Precise agriculture entails knowing the crop via the use of information technology technologies. Another strategy for predicting crop yields in advance employed time series vegetation indices to develop a crop growth process model. The approach using several indices was demonstrated to be slower than the one using a single index. Using (ANN) artificial neural networks, our study predicts crop production depending on location, temperature, and other factors. So that farmers can figure out when the best time is to plant certain crops. This boosts their earnings as well. Data is gathered and converted to a specific format for data transmission. Then, using the ANN algorithm, it is trained (classified) and thus the results can be achieved. By taking into account the state, temperature, and other variables, the result displays which crop may produce the highest yield. The crop yield will be forecasted in kg/hector depending on the particular crop, year, and region. The weather data may be shown in several ways, such as high temperature, low temperature, cloud cover, and so on. Graphs will also be presented depending on the conclusion.

REFERENCES

[1] Lobell, David B., and Christopher B. Field. "Global scale climate–crop yield relationships and the impacts of recent warming." Environmental research letters 2.1 (2007): 014002.
[2] Sakamoto, Toshihiro, Anatoly A. Gitelson, and Timothy J. Arkebauer. "MODIS-based corn grain yield estimation model incorporating crop phenology information." Remote Sensing of Environment 131 (2013): 215-231.
[3] Su, Tao, Shao Yuan Feng, and Xing Yuan Cui. "Regional Yield Estimation for Spring Maize with Multi-TemporalRemotely Sensed Data in Jinchuan, China." Advanced Materials Research. Vol. 610. Trans Tech Publications, 2013.
[4] Thomas van Klompenburg. “Crop yield prediction using machine learning: A systematic literature review” (2020).
[5] Zhang, Xiying, et al. "Contribution of cultivar, fertilizer and weather to yield variation of winter wheat over three decades: A case study in the North China Plain." European Journal of Agronomy 50 (2013): 52-59.
[6] P. Mohan and K. Patil, “Weather and Crop Prediction Using Modified Self Organizing Map for Mysore Region,” International Journal of Intelligent Engineering and Systems, vol. 11, no. 2. The Intelligent Networks and Systems Society, pp. 192–199, Apr. 30, 2018. doi: 10.22266/ijies2018.0430.21.
[7] A. Mahato, “Climate Change and its Impact on Agriculture”, International Journal of Scientific and Research Publications, Vol.4, No.4, pp. 1-6, 2014.
[8] J.L. Hatfield, K.J. Boote, B.A. Kimball, L.H. Ziska, and R.C. Izaurralde, “Climate impacts on agriculture: implications for crop production”, Agronomy Journal, Vol.103, No.2, pp.351-370, 2011.
[9] T. Mavromatis, “Spatial resolution effects on crop yield forecasts: An application to rainfed wheat yield in north Greece with CERES-Wheat”, Agricultural Systems, Vol.143, pp.38-48, 2016.
[10] L. Hong-ying, H. Yan-lin, Z. Yong-juan, and Z. Hui-ming, “Crop yield forecasted model based on time series techniques”, Journal of Northeast Agricultural University (English Edition), Vol.19, No.1, pp.73-77, 2012.
[11] P. Roudier, B. Muller, P. d’Aquino, C. Roncoli, M.A. Soumaré, L. Batté, and B. Sultan, “The role of climate forecasts in smallholder agriculture: lessons from participatory research in two communities in Senegal”, Climate Risk Management, Vol.2, pp.42-55, 2014.
[12] Supreeth S, & Shobha Biradar. (2013). Scheduling Virtual Machines for Load balancing in Cloud Computing Platform. International Journal of Science and Research (IJSR), 2(6, June 2013), 437–441. https://doi.org/10.5281/zenodo.6423763.
[13] S. Supreeth and K. K. Patil, “Virtual machine scheduling strategies in cloud computing- A review,” Int. J. Emerg. Technol., vol. 10, no. 3, pp. 181–188, 2019. https://doi.org/10.5281/zenodo.6144561.
[14] Supreeth S., & Kirankumari Patil, “Hybrid Genetic Algorithm and Modified-Particle Swarm Optimization Algorithm (GA-MPSO) for Predicting Scheduling Virtual Machines in Educational Cloud Platforms”, International Journal of Emerging Technologies in Learning (iJET), 17(07), pp. 208–225, 2022, https://doi.org/10.3991/ijet.v17i07.29223.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)