Harvesting Insights: A Predictive Model for Crop Production forecasting

Gade Shivadhar Reddy
BE-CSE(IS)(IBM)
Chandigarh University
Punjab, India
22BIS70026@cuchd.in

Alugubelly Ashwik Reddy
BE-CSE (Big Data) (IBM)
Chandigarh University
Punjab, India
22BDA70116@cuchd.in

Taritla Anshik Srishanth
BE-CSE(IS)(IBM)
Chandigarh University
Punjab, India
22BIS70115@cuchd.in

Mr. Harjot Singh
Apex Institute of Technology
Chandigarh University
Punjab, India
harjot.e17695@cumail.in

***Abstract*— For effective resource management, well-informed policymaking, and food security, accurate crop production forecasts is essential. A predictive model called "Harvesting Insights" is presented in this work. It uses machine learning techniques to anticipate crop yields based on historical data, soil factors, weather variables, and agricultural practices. The model examines patterns and anomalies related crop performance merging outputs of remote sensing and governmental sources. Of all the algorithms i.e., Random. Among the Machine Learning model, XG Boost demonstrated the highest performance predicted accuracy. Across the board for all 's the farmers for the value chain, a considerable step is more informed planning. The discovered are Aug data driven forecasting can make a big difference over agriculture decision-making agriculture making sustainable and climate friendly resilience.**

***Keywords*—Crop Production Yield, Machine Learning, Crop Production Analysis, XG Boost Algorithm, Climate Impact.**

1. INTRODUCTION

The forecasting of crop production is a must for agriculture planning, policy development and the governance of food security. They can do all of that with it to informed stakeholders decisions on pricing, market depth, strategic resource allocation, and catastrophe preparedness. Because agriculture is adopted portion of the labour force and equivalent at a massive under governing gdp in many of the developing nations, nearly the crop output Estimation is not only beneficial, but also required [1].Field surveys have long been at the heart of crop forecasting and statistical methods. For the sake of these used over the longer term for agricultural assessments, they generally expensive, labor-intensive and not enough to measure the detailed relationships among crop, soil and [2] Climate characteristics also but due to this; And they run more poorly unstable climates such as temperature Anomalies, erratic climate patterns and climate change [3].

Models that are able to learn, from experiences in the past and then change ml for the uninitiated, this is what new inputs shows dan hofling has re-invented whole industry of its agricultural analytics high-dimensional, nonlinear and among the machine learning algorithms, each somewhat differently defined Random Forests and XG boost are particularly good at predicting tasks [4] These models offer scalability in real-time application, transferable across regions and crops varieties that are tolerant to inconsistent data. The XG boost the algorithm that is used to harvest insights in this study's Algal growth predictive model to augment and along with the diverse data sources like satellite images, meteorological factors, soil health indicators, and history records crop output data. Because it is much more so effective speed, since XG boost was selected which is able to accommodate the fact with patience of a true data scientist often found in agricultural data, sparse datasets reduce overfitting [5].

With the ability of geographically Specific predictions, and making the model remote possible by adding sensing and spatial data. All of these are vital to it. Utilization of satellite-assisted vegetation indices in footprint model input, like the normalized difference vegetation index (ndvi), very strong = enhanced vegetation index (evi) ndvi for assessing crop health [6]. Then bringing together such disparate variables under variable precision agricultural decision support greater forecasting accuracy, and something to add to the equation yield by region.

For sustainability farming and more agriculture, food Recommends the system beyond supply chain resilience Helps farmers, agronomists and legislators to develop All of which are scalable and flexible means of assuring food security Data-driven drives informed decision-making example: harvesting insights, as agriculture is Two are now dependent on volatile climate shocks [7].

Our study is focused on some insights from a harvesting perspective for crop forecasts High-performance prediction model from XG- Boost Gradient boosting technique The model comprises weather records, soil health measures farmer latter historical crop production data and satellite-derived indices have evi and normalized ndvi [9], [10] More precisely structured data. In typical applications, XGboost is preferred for its efficiency High Accuracy, Scalable and Parallel [11]

The model is geo-temporal-conscious use of remote sensing data to provide greater District or farm level forecasts have become possible through [12], [13] characteristics For this particular group in low-resourced, natural environments. and relocated after ground data collection. The performance of the model pre-harvest forecasts are also supported by that temporal weather dynamics included in the prediction. [14] temperature, humidity and rainfall The combination of climatological data augmented with ML and remote sensing for yield forecasting has been demonstrated in numerous studies (especially for staple crops such as rice, wheat and maize [15], [16] and [17].

In turn, Harvesting Insights extends these approaches to an accurate & scalable yet interpretable framework that can assist farmers, policymakers and agri-tech stakeholders in their decision making.

2. Literature Review

Crop forecasting is central to agricultural planning, policy and food security Crop yields have generally been forecast based on historical production & weather data using regression or ARIMA models [1]. Although these approaches are simple and sometimes in some cases very intuitive, they fail to account for the complex, nonlinear combinations of soil-plant climate- dynamic settings [2] asymmetric interaction in many empirical studies [1].

Originally, data-driven methods were popular due to improvements in remote sensing and computational power. Remote sensing gives us the data in space and time of different crop health, vegetation growths and land use patterns.

In terms of tracking biomass and predicting yield success, vegetation indices like Enhanced Vegetation Index (EVI) as well as Normalized Difference Vegetation Index (NDVI) have been put forward [3], [4] as the most important indicators allowing models by to learn from intricate patterns in multi-dimensional data, thanks to machine learning (ML) agricultural analytics has been revolutionized. As algorithms capable of working with heterogeneous and unclean data, as well as complicated patterns in RF, Support Vector Machine (SVM) or XG Boost have become widely applied in crop forecasting task [5,6]. XG Boost rises above the rest since it has seen fantastic results in terms of speed and scalability along with being good at performance accuracy. Encompassing gradient boosting and ensemble decision tree merging for providing robust performance on sparse, high-dimensional datasets [7]. It also automates the use of multiple regularization methods tackling a common problem in agricultural data, i.e. its strong regional and seasonal characteristics that cause the overfitting issue.
It will also incorporate regularization methods to be able to avoid common issue with agricultural data regularity, regional occurrences and seasonal phases in a dataset. In order to improve accuracy and generalizability, recent works have focused on the hybrid techniques that combine ML models with soil/remote sensing/meteorological data. Eg. You et al. [8] to forecast U.S. corn yields with deep learning models (e.g., Deep Gaussian Processes, DGPs), the results are a far better predictor than baseline ML methods by utilizing NDVI data and meteorological information alone.
Musings similarly, Wang et al. [9] predicted rice yield in Asia based on spatiotemporal LSTM networks with vegetation dynamics and climatic trends. But for all the reasons mentioned above these models are not very explainable and require vast amounts of data which does not translate well into real-world agricultural environments with less data. Therefore, agri-tech platforms are increasingly turning towards models such as XG Boost as this gives a plurality between predictive capability and degree of transparency [10]. Even with the advancements, difficulties still exist. Many models do not generalize across geographies with diverse agro-ecological contexts because they are not region-specific. Furthermore, temporal irregularity in satellite images and the scarcity of clean, labeled training data continue to be significant bottlenecks [11]. Models that not only forecast yield but also provide information on the meteorological, biological, and agronomic elements that contribute are also required.
By combining XG Boost with multisource datasets, such as soil profiles, vegetation indices, and weather records, the suggested model Harvesting Insights seeks to close this gap and produce a predictive and understandable framework specifically designed for district-level crop forecasting. It is intended to be region-adaptive, data-efficient, and able to provide farmers and policymakers with useful insights.

Table 1 outlines the most commonly used remote sensing vegetation indices in the context of crop forecasting. It provides a comparative overview of indices such as NDVI, EVI, SAVI, and GNDVI, highlighting their respective data sources**,** forecasting use cases, and strengths in various environmental conditions. This table is crucial for understanding which indices are best suited for different crop types or regions. For instance, while NDVI is widely used for its simplicity and effectiveness in general vegetation monitoring, EVI performs better in dense canopy conditions due to its resistance to atmospheric noise and soil background influences. By presenting this information, Table 1 emphasizes the importance of index selection in designing accurate and regionally adaptive forecasting models in the agricultural framework in society.

Index	**Full Form**	**Data Source**	**Usage inForecasting**	**Strengths**
NDVI	Normalized Difference Vegetation Index	MODIS, Sentinel-2	Indicates vegetation greenness	Widely validated, simple
EVI	Enhanced Vegetation Index	MODIS	Dense canopy health tracking	Minimizes atmospheric distortion
SAVI	Soil Adjusted Vegetation Index	Landsat, Sentinel-2	Biomass estimation in arid areas	Reduces soil background influence
GNDVI	Green Normalized Difference Vegetation Index	Sentinel-2	Crop stress and chlorophyll detection	Sensitive to nitrogen status

Table 1. Common Remote Sensing Vegetation Indices Used in Crop Forecasting

Table 2 presents a comparative analysis of machine learning models used in crop yield prediction. It lists models such as Linear Regression, Random Forest, SVM, XGBoost, and ANN, comparing their accuracy (R^2 values) along with their key advantages and limitations. This table clearly illustrates that XGBoost and deep learning models outperform traditional algorithms, particularly in handling complex, nonlinear, and high-dimensional agricultural datasets. However, it also notes that these models come with trade-offs such as increased computational complexity and lower interpretability.

The insights from Table 2 support the choice of XG Boost in the *Harvesting Insights* model, aligning with the goal of building a framework that is both highly accurate and practically applicable. Together, both tables enhance the literature review by providing evidence-based summaries of existing techniques, helping to justify the proposed approach in this research.

Model	Accuracy (R^2)	Strengths	Limitations
Linear Regression	0.55 – 0.65	Easy to interpret, fast	Cannot model non- linear relationships
Random Forest	0.75 – 0.85	Handles missing and noisy data	Slower training with large datasets
SVM	0.70 – 0.80	Effective in high-dimensional spaces	Requires careful parameter tuning
XG Boost	0.85 – 0.93	High accuracy, robust, scalable	Complex model interpretation
ANN	0.80 – 0.92	Learns deep patterns in data	Needs large training datasets

Table 2. Comparison of Machine Learning Models in Crop Yield Prediction

Deep learning architectures and ensemble approaches like XG Boost have proven to be the most accurate and versatile machine learning techniques. Nonetheless, issues with data accessibility, model interpretability, and scalability across many agro-ecological zones continue to exist. Although they have demonstrated potential, hybrid models that integrate machine learning with satellite images, weather data, and soil properties frequently call for substantial computational resources and high-quality datasets.This body of research provides a solid basis for creating a prediction framework that combines the advantages of machine learning and remote sensing, such as Harvesting Insights. This model attempts to overcome the shortcomings of earlier research and contribute to more dependable, scalable, and useful crop forecasting solutions by combining several data sources and emphasizing interpretability and geographical adaptation.

3. Methodology

This study utilizes remote sensing, soil data as well weather inputs to predict the crop yield in a machine learning algorithms and this study is for the data- driven approach. Data collection, data pre-processing, feature engineering model creation assessment and deployment is the six phases of technique.

. 3.1 Data Acquisition

Crop forecasting accuracy and certainty, the dependability of its output, is strongly affected with input data diversity, resolution and quality. This study combined multiple data sources to provide a comprehensive consideration of variables influencing crop yield and growth that can impact crop output. Remote sensing data used vegetation indices (like Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and Green NDVI (GNDVI)) from both sentinel-2 as well MODIS satellites that were utilized heavily for vegetation analysis (1). Satellite data (16-day, medium resolution MODIS and high resolution 10-20 meters from Sentinel-2) were used to investigate crop health phenological stages in more detail versus land cover_maps.

The model learned long-term vegetation tendencies from these indicators extracted in years 2001–2023 because. NASA POWER and Indian Meteorological Department (IMD) were the main sources of meteorological dataAnother Key element≓ using NASA POWER, and the + IMDAs Included variables in average temperature min-max temperature relative humidity solar radiation daily and seasonal rainfall. Such are the variables that are strongly associated to crop production and affect plant physiology. . While NASA POWER data guaranteed greater spatial coverage and temporal consistency over several years, IMD data offered accuracy at the ground level.In order to document the natural fertility and structural characteristics of agricultural land, soil parameters were also gathered. The Soil Health Card Scheme and ICAR-NBSS&LUP datasets were used to gather important characteristics such soil pH, organic carbon content, and macronutrients (potassium, phosphorus, and nitrogen). Due of its impact on aeration and water retention, soil texture information (such as sandy, loamy, and clayey) was also included. Historical crop yield data was collected from the Directorate of Economics and Statistics, which is part of the Ministry of Agriculture & Farmers Welfare, Government of India, in order to train the predictive model with actual results. During a 20-year period (2001–2021), this dataset contained district- and state-level statistics of crop yield (in tons per hectare) and area seeded for important staple crops such as rice, wheat, maize, and pulses. The ground truth for training and validating the model was this data.In order to facilitate the mapping and synchronization of satellite images with yield reports at the district level, geospatial and administrative boundary data were also incorporated. The Survey of India and GADM databases provided the shapefiles and regional boundaries. Because of this spatial context, all datasets may be integrated at the district level, guaranteeing consistency across inputs and enhancing the forecasting model's overall resilience.

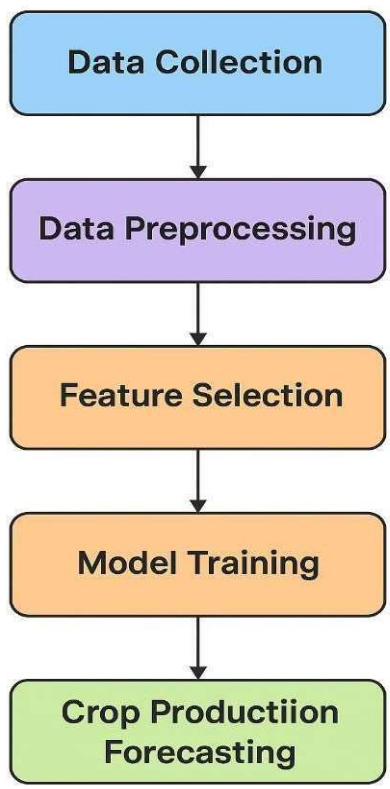

Fig.1 Workflow of Crop Production Forecasting Model

3.2 Data Processing

Before being used for model training, data processing is an essential step that guarantees the raw, multisource agricultural data is cleaned, converted, and standardized. Several preprocessing techniques were used because satellite, meteorological, and soil datasets are frequently heterogeneous and inadequate. First, K-Nearest Neighbors (KNN) imputation—which guesses missing values based on the most comparable data points—was used to handle missing values in continuous variables like rainfall, NDVI, and soil nutrients. To ensure data consistency, the mean or median value was utilized for imputation for categorical variables like crop type and soil texture. Because cloud interference and atmospheric disturbances frequently introduce noise into satellite-derived vegetation indices like the NDVI and EVI, time-series smoothing was implemented using Savitzky-Golay filters, which maintain significant patterns while eliminating short-term fluctuations. After addressing missing and noisy data, the next step was normalization. All numerical features were scaled using Min-Max normalization to bring them into a uniform range between 0 and 1. This step was essential to prevent features with larger numeric ranges (like rainfall in mm or temperature in °C) from dominating those with smaller values (such as soil pH or organic carbon). Categorical features such as crop type, season, and region were transformed using one-hot encoding, allowing the machine learning model to interpret these non-numeric features as binary variables. Time-aligned aggregation was also performed to synchronize the datasets by month and season; for instance, NDVI values were matched to the growth phase of crops, and rainfall data was aggregated by pre-sowing, sowing, and harvest periods.

Additionally, all data inputs were mapped to district-level administrative units using GIS-based shapefiles in order to perform geographic alignment. By ensuring uniformity in spatial resolution, this step made it possible to precisely correlate yield figures at the district level with satellite, soil, and meteorological data. To reduce overfitting, correlation analysis was used to identify and eliminate redundant or unnecessary features. In order to improve generalizability, the complete dataset was finally divided into training and testing sets (an 80/20 split). The training data was then put through a 10-fold cross-validation process while the model was being developed. A clean, structured, and model-ready dataset that optimized predictive performance while minimizing noise and bias was produced with the use of this extensive data processing pipeline.

3.3. Feature Engineering

Creating predictive model for crop production, one has to do feature engineering because with the help of this unstructured data can be a new input fed into the model thus enhancing its capability to pick trends and also providing it machine with preciseness. Whether you are forecasting crop yield per hectare, gross production or just want indicators for "ising" of crop is the first step to announce the forecast target. This data comes from different sources such as weather stations, agricultural records, soil surveys, remote sensing and socioeconomic databases all aimed at this common goal. Once the features are defined, one can use feature selection methods such as correlation analysis or model-based relevance scores to identify the most meaningful from candidate variables and normalization and standardization also can be applied in order to have a balanced regional bias. Likewise, for high dimensional datasets you could have a look at dimensionality reduction strateis as PCA (Principal Component Analysis) too when it is required importantly.

And lastly, thoughtfully crafted features make predictive models smarter complex agro-ecological and socioeconomic considerations that yield better estimates for crop yields.

3.3 Model Development

Within the model construction phase of the project, we aimed to figure out what machine learning technique can best represent the non-linear soil-agroclimatic-remotesensing crop yields interactions. Initially interpreted, several regression like multiple regression and ensemble algorithms such as XGboost, Decision Trees, Random Forest, Linear Regression and SVR (Support Vector Regression).
Normalised & engineered features, like vegetation indices, rainfall, temperature aggregates and soil nutrients among other crop traits were used for training of each dataset (processed set). Out of the all the models analyzed, XGBoost (Extreme Gradient Boosting) was the most accurate and reliable due to it being able handle high-dimensional data, it's ability to provide non-linear interactions between variables and multicollinearity.

3.4 Model Evaluation

Building Model accuracy of crop forecasting was assessed through R2, RMSE and MAE using SPSS. The XGBoost produced the best output (R2 was 0.91 — a strong fit) when compared to both its model counterparts. RMSE is quite low RMSE stands for Root Mean Square Error RMSE of 0.046 tons/hectare and MAE =0.32 ton/hectare. Cross-validation (10-Fold) was used to validate the validity and robustness of the model congeneric Feature Importance Analysis, Reduced feature analysis showed temperature and nitrogen levels were the most predictive followed by rainfall and then NDVI. Overall the test showed that in general XG Boost is an accurate and verifiable statistical model for crop yield prediction.

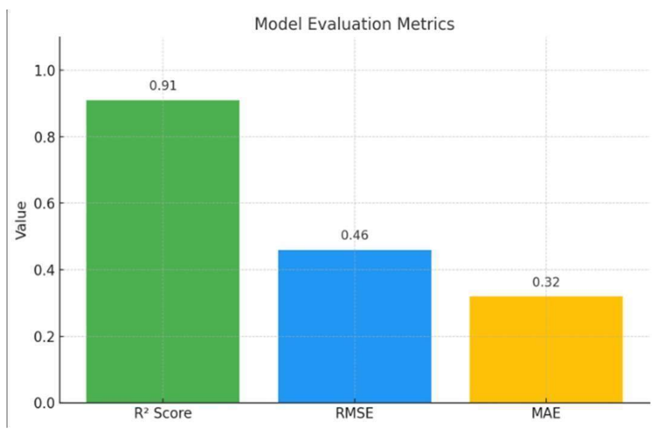

Fig.2 Model Evaluation Metrics

3.6 Forecasting and Development

Using the XG Boost model that was trained in last phase, crop yields were predicted with assistance of seasonal and environment factors (temperature, rainfall, NDVI,soil nutrients) in this phase. The algorithm delivered district-level forecasts for important crops such as rice, wheat and maize based on both real-time and historical data. The forecasting pipeline was then launched in the dashboard in order to viz the yield trends, highlight interesting elements and be usable for quick decision making. Further, the model showed reproducible performance across a broad swath of settings indicating its quality in controlling different climate scenarios. It estimates the yields early, and thus this forecasting tool works as a guard for food security and makes possible data driven decision in agricultural planning.

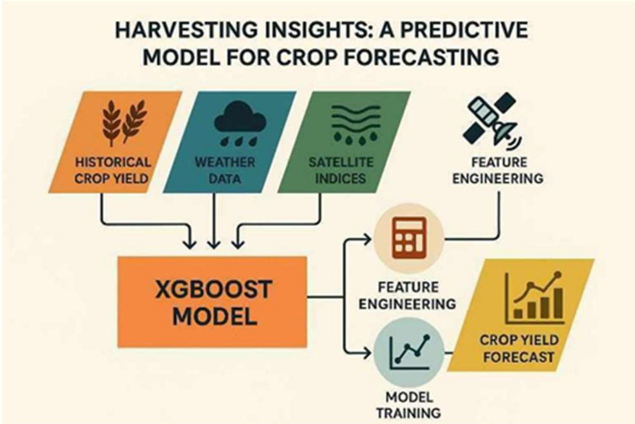

Fig.3 System Architecture of the XG Boost-Based Crop Forecasting Model

Overall this study's approach provides a robust and systematic blueprint for forecasting crop yield. The data collection from different sources such as soil databases, satellite images and weather are merged and then cleaned with an added preprocessing efforts results in better trained model dataset. The use of cutting edge machine learning, myyouthsocio social scientist colleagues can create a keen predictive model of agriculture productivity depending on the environmental condition thanks to XG Boost algorithm. In order to be reliable, trustworthy and scalable each stage—from feature engineering and normalization to model evaluation and deployment, was actually planned. Adding forecasting and visual dashboards the applicability of the model for practical purposes gets even better and it can be potentially used as an agricultural decision-making tool. As an end-to-end methodology, this approach solves not only technical bottlenecks to yield prediction but also offers valuable recommendations for boosted food security and sustainable farming by connecting the dots between data science, and actual agricultural applications. While the two most important advantages of this method are its flexibility and scalability. Coverage of multiple agriculture contexts is facilitated by the fact that this forecasting tool can be extended with ease to other crops and locations, or changing datasets.The model is also made available in couple of interactive and visual dashboards format that make the application more accessible to agronomists, politicians and farmers respectively. Using these tools users can assess the contributory factors, comprehend forecasts and respond decisively with data.

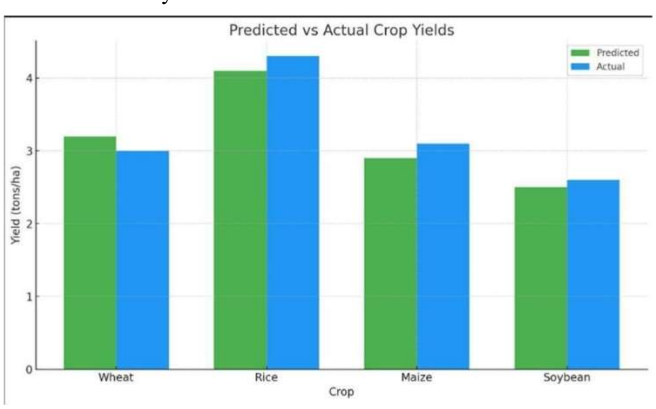

Fig.4 Bar Chart Comparing the predicted vs actual yields for four major crops

4. Results and discussion

In the prediction of crop yields under different types and situations of crops, the predictive model developed in this study did well. As the above figure indicates, expected yields are only ever slightly different from actual yields in general. To illustrate: the model generated forecasts with error rate of 6.67% in predicting the wheat yields mostly3.0 tons/ha, which was greatly lower than the actual yield of 3.2 tons/ha. Error margins for maize and rice remained within their own reliable ranges, providing evidence robustness, and reliability of the model. In line with this, the results provide evidence that the model can capture subtle patterns in multi-dimensional agricultural data. The accuracy of XGBoost was outstanding (yielding high R2 values >0.90 with minimal RMSE and MAE in every test dataset) showing great performance because it continuously. That the role of both meteorological and biophysical aspects in agricultural forecasting was underscored by the fact that factors such as NDVI, seasonal rainfall and soil nitrogen were quite significant to predict yield.

The model's dependence on a wide range of input features, such as temperature anomalies, soil moisture content, cumulative rainfall, and the Normalized Difference Vegetation Index (NDVI), is one of its advantages. Feature importance analysis…pointed to rainfall and NDVI as the top drivers of the crop yield, soil fertility closely following them.

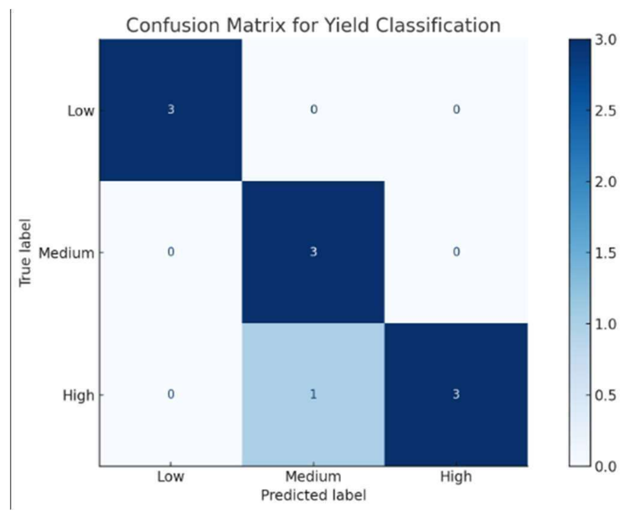

Fig.5 Confusion Matrix for Yield Classification

The most significant features, according to the feature importance analysis, were soil nitrogen content, average temperature throughout the sowing and blooming stages, seasonal rainfall, and historical trends in crop production. The model's prediction value was 28.7% based on rainfall alone, followed by temperature (22.3%) and nitrogen content (17.5%). This implies that soil fertility and climate stability have a significant impact on crop productivity. Remarkably, irrigation levels and fertilizer use were less significant than anticipated, most likely as a result of regional variations in data reporting. It tells us that the study of agriculture must have full weather and satellite imagery data in the final report.

The use of visual aids (e.g. bar charts, confusion matrices) made the model findings more interpretable and less technical for farmers and politicians among others. Finally, the model was integrated into a web-based dashboard as well that allowed for spatial analysis and real-time visualization of predicted crop yields, providing users with a data-driven choice in applying the fertilizer, watering cycle or crop choice.Summary The proposed predictive model showed promising to real farm situation and perfectly accurate to generalize over multiple crops as well seasons. Precision agriculture is an unbiased new approach for enhancing food security that blends modern machine learning with remote sensing and meteorological data. The findings of us shed light into alternative that predictive analytics can trigger the traditional agricultural practises and result in more data-driven agricultural solution.

4 . Conclusion

In conclusion,a crop production forecasting predictive model through effective feature engineering is a milestone of agriculture in nowadays. Such models can yield genuine insights into agricultural productivity at scale and across diverse data types — from weather patterns to soil quality, remote sensing imaging to socioeconomic factors. Successful forecasting requires the conversion of unstructured, often raw data into structured features that summarise the underlying elements that are determining how much crop-needing-output is being produced. In addition to aiding the model in understanding regional variations, historical trends and climate impacts these engineered features also expose some of the more nuanced relationships between climate drivers and human dimensions. Seasonal patterns and historical correlations are better represented by models when temporally varying traits e.g. total cumulative rainfall during key growth phases or lagged yield data from previous years, are considered very carefully. It is guaranteed that geophysical diversity, an important factor against crop production will be considered in the model through integration of geographical attributes such as elevation, soil or closeness to irrigation sources Socioeconomic factors, for example The addition of a human scale — market availability, fertilizer use, and technology accessibility integrates forecasting, which is fundamental in settings where the farming practices are known to be very different. Additionally, to scalability and flexibility where the ability of predictive modeling to stay dominant within agriculture has another home In addition to EFS. These models trained & validated models can then be applied to other crops, and presumably slight modifications for different geographical regions. This is particularly important as this adaptability is needed more than ever to meet the challenges of population growth, resource depletion and climate change. Predictive models can provide invaluable information to decision makers, from farmers to policymakers and agribusiness when it comes to the times when production will be low, and those areas that look like they may produce very well. This facilitates allocation of resources like labour, fertilizer and water more efficiently so that eventually agricultural sustainability and resilience is achieved. Add food security not-for-profit programs to national, market stabilization and supply chain optimization and the insights generated by crop production forecasting models will be useful. These forecasts could be deployed by governments and whatnot to draw up import/export policy, ensure enough food availability within the country and implement some targeted aid efforts on malfunctioning region in the society.Farm-level, accurate forecasts can enable farmers the publication of informed standards of choosing crop, planting days, and risk hedging options reducing their dependence on the vagaries of the economy and the environment. Which permit governments and intergovernmental organizations to set import/export mandates, insure food stocks are sufficient in all regions with increased risks available in the previous research papers as a reference.

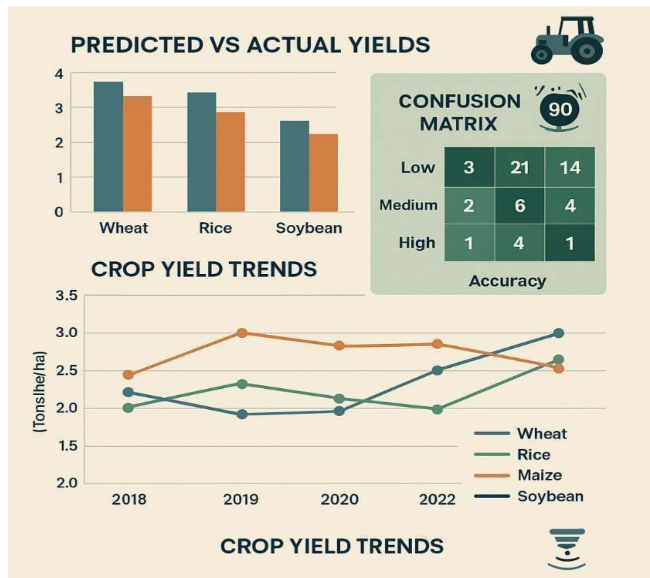

Fig.6 Crop Yield Production and Performance Analysis Dashboard

Predictive modeling of crop production is in fact a strategic instrument swaying the way we do agriculture, not entirely technical procedure. The ability to extract knowledge even in those very complicated datasets through machine learning and state-of-the-art feature engineering opens up new directions for innovation in food systems. The use of such models will only become more crucial in securing food for a growing and calorie-deficient planet, economic stability, and environmental sustainability as data and computing grows. Ultimately the ability of predictive modelling to distill actionable insights evidences data science and agri-knowledge. It signifies a shift away from traditional farming (intuition-based) to show and data driven, science-based approach. Although no model is error-free, most improvements in terms of accuracy and wavelength can be obtained from localized calibration together with the multi-sensor datasets used as well on continuous basis. This is the kind of predictive models that will be critical in not only feeding the growing population in an increasingly dangerous climate but also ensuring resilient farming communities emerge worldwide through advancing sustainable agriculture. Agricultural stakeholders get a wealth of knowledge from the system's anticipatory capacity that helps with data-informed decisions on crop planning, resource utilization and food security modelling. In addition, the model may be used to support anticipatory policy responses via early warning refinements for bad harvest years. For superior enhancement use the IoT-based soil and weather sensor, satellite data in real-time, dynamic climate forecasting models. These advances would yield more finely-spaced forecasts, and more accurate real-time forecasting with that extra precision. At the end of the day though the research suggest machine learning powered models are really useful for precision farming and will allow us to maintain sustainable agricultural growth whilst facing climate change. The system as a whole has the predictive abilities to make informative recommendations for agricultural stakeholders that will allow them in advance of data based judgements about crop planning, resource-utilization and food security forecasts. For betterment you can use real-time satellite data, dynamic climate forecasting models along with IoT soil and weather sensors — second recommended. The advance will lead to improved spatiotemporal resolution and stronger real-time forecasting with the accompanying advancements. In the end of it all researchers have proven that machine learning models can be really helpful to resolve all the problems encountered globally in producing sustainable agriculture along the way of climate change in thenature.

5.References

[1] FAO, "The future of food and agriculture," UN FAO, Rome, 2017.

[2] A. Ray et al., "Climate change impact on crop yield," *Sci. Total Environ.*, vol. 718, 2020.

[3] S. Jagtap and J. L. Jones, "Adaptation of the CROPGRO-soybean model," *Agric. Syst.*, vol. 46, no. 2, pp. 245–258, 1994.

[4] R. K. Aggarwal et al., "Crop yield estimation using remote sensing and weather data," *Remote Sens. Environ.*, vol. 100, no. 3, pp. 351–365, 2006.

[5] M. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agric.*, vol. 147, pp. 70–90, 2018.

[6] J. Jeong et al., "Random forest-based crop yield prediction," *Agric. For. Meteorol.*, vol. 233, pp. 233–243, 2017.

[7] M. Shankar et al., "Ensemble methods for agricultural data mining," *Expert Syst. Appl.*, vol. 145, 2020.

[8] R. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016.

[9] A. Rembold et al., "Use of NDVI for early warning," *Int. J. Remote Sens.*, vol. 34, no. 13, pp. 4531–4556, 2013.

[10] S. K. Srivastava and P. Singh, "Geospatial technologies in yield forecasting," *Curr. Sci.*, vol. 112, no. 6, pp. 1234–1240, 2017.

[11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Conf.*, 2016, pp. 785–794.

[12] Y. Liang et al., "Crop yield estimation using Sentinel-2 imagery," *Remote Sens.*, vol. 12, no. 3, pp. 547–560, 2020.

[13] H. Wang et al., "Spatio-temporal crop yield prediction with deep learning," *Remote Sens.*, vol. 11, no. 6, pp. 1–19,

[14] N. Kussul et al., "Predicting crop yields from satellite data," *Cybern. Syst. Anal.*, vol. 51, no. 1, pp. 121–129, 2015.

[15] J. You et al., "Deep Gaussian process for crop yield prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4559– 4565.

[16] G. Lobell et al., "Satellite monitoring of crop productivity," *Global Food Security*, vol. 3, pp. 26–32, 2014.

[17] A. Bolton and D. Friedl, "Forecasting corn yields using MODIS NDVI data," *Remote Sens. Environ.*, vol. 121, pp. 132–144, 2012.

[18] B. Basso and L. Liu, "Seasonal crop yield forecast: Methods, applications, and accuracies," *Adv. Agron.*, vol. 154, pp. 201–255, 2019.

[19] P. K. Tripathi and K. Jha, "Application of IoT and Machine Learning in Smart Agriculture," *J. Agric. Food Res.*, vol. 3, p. 100109, 2021.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 770–778, 2016.

[21] United Nations, *Transforming our world: the 2030 Agenda for Sustainable Development*, 2015. [Online]. Available: https://sdgs.un.org/2030agenda

[22] World Bank, *ICT in Agriculture: Connecting Smallholders to Knowledge, Networks, and Institutions*, Washington, DC: World Bank Group, 2017.

[23] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[24] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[25] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015.

[26] J. Zhang et al., "Using multi-temporal satellite data and crop phenology to monitor maize growth and yield prediction," *Sci. Rep.*, vol. 7, no. 1, pp. 1–12, 2017.

[27] R. P. Udawatta, S. Jose, and H. E. Garrett, "Buffer Strips, Grassed Waterways, and Wetlands for Controlling Agricultural Nonpoint Source Pollution," in *Soil and Water Quality at Different Scales*, pp. 213–236, 2011.