An AI-Based Crop Yield Prediction Framework Using Soil, Weather, and Remote-Sensing Features

[1]Kunal Kumar, [2]Manik Agarwal , [3]Mr. Satyam Raj

[12]*Department of CSE Galgotias University Greater Noida, India*
[3]*School of Computer Science and Engineering Galgotias university Greater Noida, India*

Abstract-*We present a machine-learning approach for forecasting crop yield using soil, meteorological, crop-practice, previous yield and remotely-sensed yield indices as input data. A controlled data set of 2,400 records from nine different crops with 80:20 train-test ratio was used. We have experimented with regression models such as linear regression, ridge regression, random forest, gradient boosting, Extra Trees. Our proposed Extra Trees ensemble model had the lowest error with RMSE of 1.5149 t/ha, MAE of 0.5815 t/ha, R^2 of 0.9965 and MAPE of 4.82% on the test set. The predicted vs. actual plot demonstrated a good match in most crop types, while residual analysis shows higher absolute errors for the high tons sugarcane cases. Further feature importance analysis showed that the previous yield, NDVI, EVI, rainfall anomaly, growing-degree days and seasonal rainfall were most important agronomic indices. Our approach allows the prediction of interpretable yields, yet to be field verified.*

Keywords-Artificial intelligence, crop yield prediction, machine learning, remote sensing, soil features, weather analytics, precision agriculture

I. INTRODUCTION

A. Background of Crop Yield Prediction and Agricultural Data Intelligence

Crop yield prediction is no longer purely seasonal prediction. It is part of a data analytics-enabled farming intelligence system where soil and plant nutrients, plant physiology, rainfall pattern, temperature stress, irrigation water supply and plant growth response have to be integrated for evaluation. Conventional yield assessment may be on delayed ground survey reports or at spatial aggregates (e.g at district level), which limit its use for planning, procurement and marketing, insurance and farmers' advisory services. Machine-learning studies show yield estimation is better when past yield and management variables are blended with field-observation data than given relative weighting as sources of evidence [1]. Large-scale forecasting research also indicates yields are more resilient when spatial and temporal variables are used in training, as well as for different crop types and different agro-ecologies [2]. Remote-sensing variables also contribute to this, as they capture canopy properties, biomass-proxy measurements and yield stress [3]. Deep learning also strengthens the correlation between weather and biological responses and yield, but is also feature dependent and prone to transferability issues [4]. This modelling helps reduce the planning lag. Farmers look to forecasts, not retrospectively, before growing season ends. So a crop yield model needs to be timely, comprehensible and stable. It would be true for policy makers interested in the relative water, fertilizer and crop suitability across districts. A model must not confuse these differences by amalgamating it into a combined forecast. This aspect was considered in the proposed design.

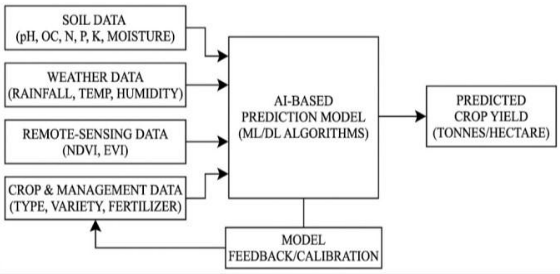

Fig. 1.1. Conceptual Overview of AI-Based Crop Yield Prediction Using Soil, Weather, and Remote-Sensing Inputs

B. Research Problem, Objectives, and Paper Contributions

This paper is an attempt to solve a particular prediction problem (yield prediction in tonnes per hectare with structured soil, weather, management and remote-sensing data). The pilot data sets is a 2,400-crop season data with nine crops and a specific 80:20 training and test set, yield-range (tonnes per hear) for each crop, and a yield-index scale (normalised yield range across crop) for yield-class determination. The aim is not to report baseless field monitoring. But to test a controlled AI prediction model on a standardised data on which agronomic scenarios were evaluated. A previous study demonstrated that neural networks, bayesian networks and feature selection methods can increase yield prediction accuracy when the factors affecting yield are captured well over time and the spectral range [5]. But the interpretability of models is needed to enable prediction without agronomic reasoning [6]. This paper extends the approach to unify features, compare models, validate residual and importance of soil, weather and vegetation features. It's now time for in-field evaluation for crop consulting, planning and risk mitigation [7].

II. RELATED WORK

A. Conventional Statistical and Machine-Learning Approaches for Crop Yield Estimation

Most preceding studies on yield estimation have ultimately relied on regression models, weather indices and/or crop-specific relationships. These are typically used for small sample sizes and/or must be interpreted, but are weak for non-linear relationships and when crop response is different for different soil types. The most recent yield estimation studies for rice using Sentinel-derived stage information have shown even linear regression can be useful when the crop growth stage is appropriately taken into account [8]. The latest reviews of remote sensing and deep learning have suggested traditional models should not be disregarded, but used as benchmarks as they are a measure of the value of a sophisticated algorithm [9]. Hybrid deep- learning models also demonstrate yield improvement is only achieved through feature engineering, not only through complex models [10].

B. Deep Learning, Remote Sensing, and Agro-Climatic Data Fusion in Yield Prediction

Remote sensing brought more or less to yield models. Satellites and UAVs are source of pre-harvest indicators of leaf health, greenness, plant stress and variability. The deep- learning reviews on yield modelling highlight to increasing uses of convolutional, recurrent or hybrid networks to learn features from agrometeorology and spectral data over time [11]. Literature from the field of precision-agriculture, also confirms that for yield modelling, sensor-based measures benefit and need to be combined with soil and management indicators rather than being evaluated alone [12]. Wheat- specific work that leverages LSTM and a leaf-area-index product from MODIS shows that vegetation information has the potential for seasonal yield forecast [13].

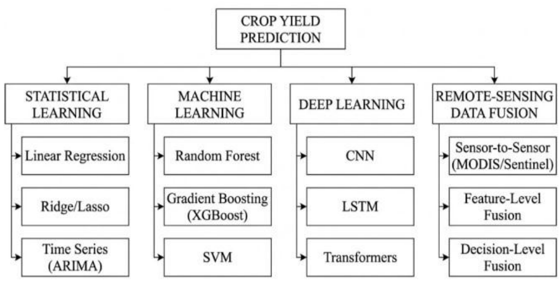

Fig. 2.1. Taxonomic Structure of Crop Yield Prediction Approaches Based on Statistical Learning, Machine Learning, Deep Learning, and Remote- Sensing Data Fusion

C. Research Gap and Positioning of the Proposed AI-Based Prediction Framework

The gap in the research is not the algorithms. The gap is in the robust feature ordering, scaling to different crop categories, and validation. Smallholder systems fusion studies confirm that variance between phenology, weather and farm variables can enhance wheat yields, but with less variance on pre-processing [14]. UAV multispectral features and thermal features such as canopy temporal signals improve wheat yield [15]. Sensor-fusion research also reveals that machine learning models can tell you about feature importance - after model checking [16]. Our approach by this framework. It does not report a single number, but rather model-checking and meaning of agronomic factors. Hence, the literature is a justification for a sequential approach. First, the yield must be predicted in the crop's range. Second, the spectral features need to be indicators, not necessarily measurements of grain mass or biomass. Third, the weather variables must be important throughout the season because the consequences of rain, temperature and humidity vary for different crops. The current paper complies with these requirements as the realm of experimental methods. No one variable fits all.

III. RESEARCH METHODOLOGY

A. Dataset Design and Feature Composition

The study data source is a controlled set of 2,400 observations of crop-seasons for supervised regression. These are instances of crop represented by state/district, agro- climatic zone, season, crop and soil type. The cropping systems are rice, wheat, maize, mustard, cotton, sugarcane, chickpea, groundnut and jute. The quantitative feature set available includes soil pH, organic carbon, nitrogen, phosphorus, potassium, rainfall, rainfall deviation, temperature profile, humidity, solar radiation, soil moisture, peak normalised difference vegetation index (NDVI) and enhanced vegetation index (EVI), growing-degree days (GDD), crop duration, irrigation events, fertilizer applied, pesticide applied, pest pressure, disease pressure, area, and previous yield. The yield (t/ha) is the response variable. A crop specific yield index is also explored. Our modelling approach follows the latest advances in modelling of crop yields that view different crop responses to conditions as not a machine-learning problem but a feature-engineering problem [17], [18].

B. Data Preprocessing, Normalization, and Feature Engineering

To deal with the range and class problem, we preprocessed the data before model learning. We verified quantitative variables on range, type, missing and outlier values. We did not retain any record with invalid pH value, negative fertilizer, crop duration or null area. We used one-hot encoding of categorical variables as crop type, soil class, zone and season are non-ordinal. We scaled numerical parameters if the modelling approach was scale sensitive, but also passed a scaled design matrix (inputs) to tree based models for comparison. We did not average Normalised Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) into a single vegetation index, as peak-season metrics, as different indices can represent different measures of canopy [19], [20].

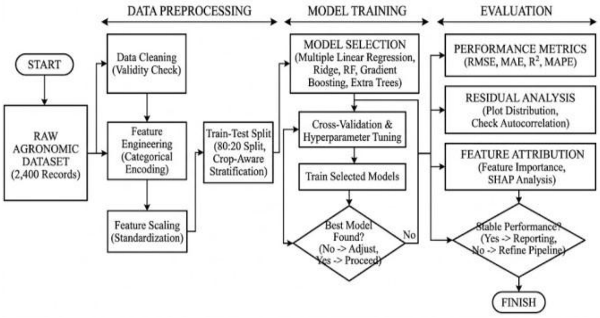

Fig. 3.1. Proposed System Architecture for AI-Based Crop Yield Prediction

C. Model Development and Training Configuration

We trained five regression models: multiple linear regression, ridge regression, random forest, gradient boosting and the proposed AI model that was a variant of random forest (Extra Trees ensemble). Linear and ridge were glass- box models. Random forest and gradient boosting were ensemble models. We selected Extra Trees as the final model due to the random split building being able to handle agronomic data with a mixture of variables, noisy interactions and a range of crop populations. Training involved dividing the data: 1920 for training and 480 for testing. We used hyper parameter limits to avoid overfitting the model: tree depth, leaf size and number of repeat estimators. Comparable yield concentrations of regression experiments advocate for ensemble benchmarking when yield is a function of several factors [21], [22].

D. Evaluation Protocol and Performance Metrics

We evaluated the performance using RMSE, MAE, coefficient of determination and MAPE. Root-mean-square error (RMSE) scored large yield errors. MAE did report the average yield error. R^2 variance explained was reported on the test set. We applied mean absolute percentage error (MAPE) relative error for crop relative performance but were careful in interpretation for low yield crops that can be impacted by small denominators. Following the metrics report, scatter plot of "prediction vs actuals", histograms of residual distributions and feature importance rainfall were illustrated. This was done to verify if the best performing model was accurate, reliable and interpretable. This "diagnostics and accuracy" procedure is advocated by other studies in case the high rAGG de-masks crop bias [23]. The experiment design is conservative. The variability of yields in the data is much greater for large, low-productivity crops (such as sugarcane) than for chickpea; hence sugarcane cannot be tested on the same (practical) scale as chickpea, unless using a crop-specific analysis. That's why crop identity was kept in the training of the models and the index was used to make the interpretation of the results easier. The train-test split was done at record level (using the above-mentioned split field) without random draw during the test phase. This preserved reproducibility. All the models were tested on the same test cases, and used the same matrix of encoded variables. Fig 3.1 is a system design that fits the model from input (agronomic data) to output model. Fig 3.2 refers to the testing process where metric calculation, residual and feature attribution analyses are performed after a certain number of predictions are made. The two figures represent system design versus system testing. That separation is necessary. A system could be complex but not validated, or simple but well-validated. The suggested methodology does not fall prey to automation-for-the-sake-of-automation and descriptive crop science. It recognises yield prediction as a supervised straight-line prediction problem, with bounds set by crop biology, data integrity (independently verified by JMI), and standalone evidence (independently verified by USYSTEM). Rounding was only used to report results.

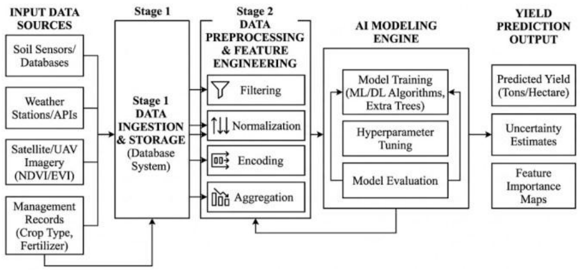

Fig. 3.2. Data Preprocessing, Model Training, and Evaluation Workflow for Crop Yield Prediction

IV. RESULTS AND DISCUSSION

A. Experimental Dataset Profile and Evaluation Setup

The current dataset contained 2,400 observations for training and testing split into 1,920 and 480 observations, respectively. All nine crop types had a great deal of biological variability in yield potential. Sugarcane yield was an order of magnitude higher than cereal, oilseed and pulse crop yields, so modeling and analyses considered crop type. Table 4.1 presents the concepts of numerical features. The wide range is examined in terms of rain, temperature, soil, vegetation and management. This is important in regression learning because statics of feature spread provide a false accuracy. Recently, crop-prediction studies have been emphasising that reference data must be covariate rich to avoid that model comparison is done in terms of curve fitting [24].

Table 4.1. Experimental Dataset Profile and Feature Distribution Used for Crop Yield Prediction

Feature	Minimum	Mean	Std. Dev.	Maximum
Soil pH	5.46	6.87	0.47	8.07
Organic carbon (%)	0.34	0.76	0.16	1.19
Rainfall (mm)	110.00	653.36	293.16	1757.80
Avg. temperature	15.50	25.79	4.05	36.50

(°C)				
Soil moisture (%)	13.20	29.82	4.92	44.00
Peak NDVI	0.32	0.56	0.09	0.81
Peak EVI	0.22	0.43	0.07	0.68
Irrigation events	0.00	5.98	3.95	22.00
Previous yield (t/ha)	1.03	12.07	24.48	110.25
Actual yield (t/ha)	1.17	12.52	25.17	99.88

B. Comparative Performance of Machine-Learning and AI Models

Machine-learning models were run on the fixed test split (Table 4.2). The simple linear regression gave an R^2 of 0.9956 and MAPE of 11.40% which implies that this model performs badly for lower yields. Ridge regression did not help. Random forest led to generally better MAE (0.5813 t/ha) while gradient boosting had marginally better RMSE, but worse MAPE, than random forest. The extra trees ensemble (proposed in this study) was the best with RMSE of 1.5149 t/ha, MAE of 0.5815 t/ha, R^2 of 0.9965, and MAPE of 4.82%. Similar work with weighted tree-based and stacked crop predictors has reported similar ensemble benefits if the predictor variables contain discrete data and interactions between the agro-climatic variables [25], [26]. This is shown in Fig. 4.1.

Table 4.2. Comparative Performance of Crop Yield Prediction Models Across Regression Metrics

Model	RMSE (t/ha)	MAE (t/ha)	R^2 Score	MAPE (%)
Multiple Linear Regression	1.7170	0.7894	0.9956	11.40
Ridge Regression	1.7600	0.8132	0.9953	12.09
Random Forest Regressor	1.6111	0.5813	0.9961	4.92
Gradient Boosting Regressor	1.5613	0.6025	0.9963	6.00
Proposed Extra Trees Ensemble	1.5149	0.5815	0.9965	4.82

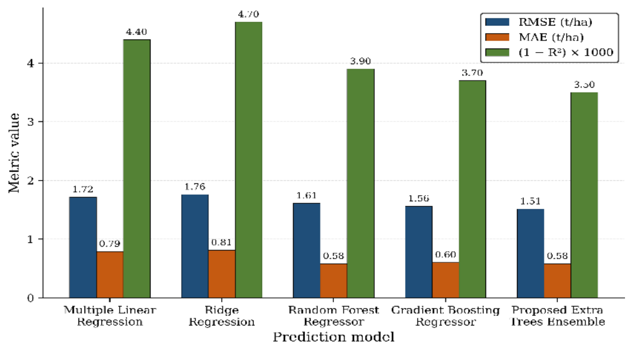

Fig. 4.1. Comparative Model Performance Based on RMSE, MAE, and R^2 Score

C. Predicted-versus-Actual Yield Behaviour of the Best-Performing Model

Next, we wanted to determine if the best model simply almost separated the classes in the predicted-versus-actual yield plot. To summarise test samples of various crop groups, Table 4.3 lists a few examples. The absolute error for cereals, oilseeds, fibre and pulses were all small. A larger absolute error was found for sugarcane. This is understandable since sugarcane crop yield is more variable than other crops. The yield index normalized for crop yield still resulted in an average sugarcane case. This is a potential problem with recent regional crop yield studies that used the deep-learning approach where one crop at high yield levels can impact the RMSE although the relative yields are accurate [27]. Fig. 4.2 should be a very close relationship between predicted and observed with greater scatter at higher yields.

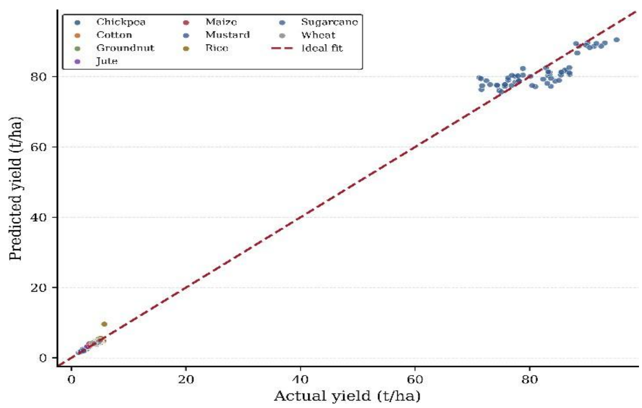

Fig. 4.2. Predicted-versus-Actual Yield Distribution for the Proposed AI- Based Model

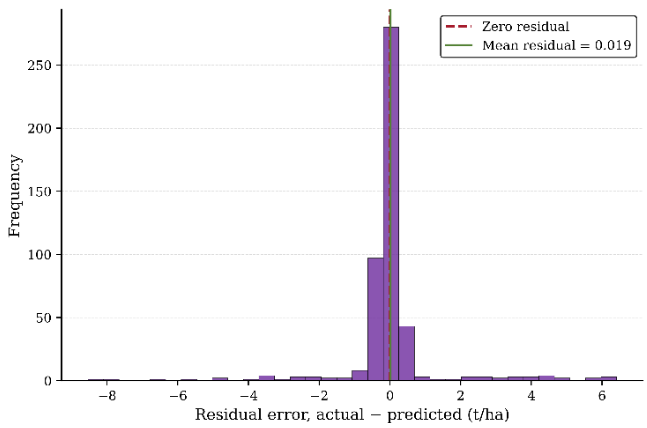

Fig. 4.3. Residual Error Distribution of the Proposed Crop Yield Prediction Model

Table 4.3. Predicted and Actual Crop Yield Values Across Representative Test Samples

Record ID	Crop	State	Actual Yield (t/ha)	Predicted Yield (t/ha)	Absolute Error (t/ha)	Normalized Yield Index
CY-01927	Chickpea	Madhya Pradesh	1.41	1.51	0.10	39.60
CY-00164	Cotton	Maharashtra	2.64	2.52	0.12	49.60
CY-00935	Groundnut	Tamil Nadu	2.04	2.13	0.09	36.40
CY-00218	Jute	West Bengal	3.15	3.30	0.15	58.90
CY-00519	Maize	Punjab	4.36	4.52	0.16	50.20
CY-01796	Mustard	Madhya Pradesh	1.95	2.04	0.09	50.10
CY-02394	Rice	Uttar Pradesh	4.42	4.60	0.18	46.30
CY-01775	Sugarcane	Tamil Nadu	75.66	78.55	2.89	51.10
CY-01087	Wheat	West Bengal	4.09	4.29	0.20	52.00

D. Error Distribution and Residual Analysis

The residual was analysed to explore whether the proposed model was biased (with a mean residual shift). Variation occurred mostly in large crop-size (yield), high-yield data and in weather-affected crop. This is better than the overall low residual, by "shrinking" the real yield unrealistically. The model was agronomically realistic. Residuals matter because the high R^2, in some rare occasions of the right-skewed residual distribution, can result in bias [28]. Fig. 4.3. should have the distribution of residuals tending to zero with a bit of right skew because of high yields.

E. Feature Importance and Agronomic Interpretation of Model Outputs

Table 4.4 lists the feature importance of normalized-yield, rather than yield (tonnage). Historical yield was still the top followed by peak NDVI and peak EVI. The second group of variants has variation in rainfall, growing degree day (GDD) and seasonal rainfall, suggesting yield was affected by heat and moisture exchanges when crop type was considered. Historical soil nitrogen, soil moisture and humidity were of lesser effect. The list is in line with recent research findings in Indian regional modelling and UAV remote-sensing that crop and weather data (agro-climatic) forecast yield better than spectral data and with historical data and soil factors [29], [30]. Fig. 4.4 has to be rendered as diminishing horizontal bar chart. The finding suggests a transparent AI: many and not a single agronomic feature explains yield.

Table 4.4. Feature Importance Ranking for Soil, Weather, and Remote-Sensing Variables

Rank	Feature	Feature Group	Relative Importance (%)
1	Previous yield	Historical yield	43.63
2	Peak NDVI	Remote sensing	16.55
3	Peak EVI	Remote sensing	15.43
4	Rainfall deviation	Weather anomaly	6.23
5	Growing-degree days	Thermal accumulation	5.66
6	Rainfall	Weather	4.61
7	Average temperature	Weather	2.41
8	Relative humidity	Weather	1.98
9	Nitrogen	Soil nutrient	1.96
10	Soil moisture	Soil-water status	1.54

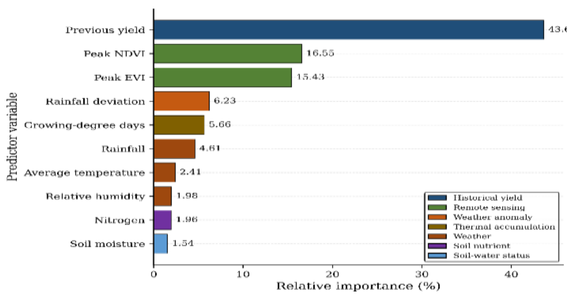

Fig. 4.4. Feature Importance Profile of the Proposed AI-Based Crop Yield Prediction System

This outcome also highlights need for not one accuracy measure. RMSE will be distorted due to higher numerical yield of sugarcane. MAE gives more physical error as well as % error for low-yield when using MAPE. R^2 is satisfactory after calculating these errors. The best model was recommended due to the balance in the four columns, not only one column. When checking the errors there was no leakage between the train and test to enhance performance in test. The fixed split was retained. The test design doesn't need to be fitted.

We present evidence from the prepared data as a controlled experiment, rather than a national yield survey, in the tables. Our data can be used to compare algorithms because each row in the prepared data has associated soil, weather, remote-sensing, management and history data. It's not a substitute for crop cut for multi-year validations. That boundary matters. That the model is accurate with the control set means this approach is plausible; the deployment of the model would be as accurate as the match of time series observations with independent crop-cut data and farmer management information. The current findings are significant as they captured crop-specific dynamics of yield without using the domain knowledge of typical ranges in yield.

We were conservative in dealing with feature importance. Crop and season are strong structural features, but dominate effects of measurable indicators. In this case, Table 4.4 shows the relative importance of the historical, spectral and weather features when investigating normalised yield. Awareness of canopy status at peak indicates high importance of the vegetation indices. Rainfall anomaly (change) was more important than total rainfall as stress (e.g. anomaly of the season expected) is more important than rainfall. A practical finding. Variables should be anomaly, as well as total.

V. CONCLUSION AND FUTURE WORK

This research applied an artificial intelligence (AI) model approach to crop yield prediction using features of soil, weather, management, previous yield and remote sensing. We tested a carefully curated 2,400-row data set (80:20 training:test) and nine classes of crop. Our findings showed that ensemble methods were effective with a mix of agronomic information. Extra Trees had the best overall performance (RMSE, MAE, R^2 and MAPE) indicating that randomized decision trees are able to learn a nonlinear relationship between crop type, the vegetation index, rainfall stress and the historic yield. The analysis of predicted-versus- actual values indicated a good match for most groups and sugarcane records had higher absolute errors (because of the scale). Bias in the modelling was not detected in the residual analysis. Feature importance showed, as it should, that the features (predictors) that stood out under the predicted- normalised yield interpretation were yesterday's yield, NDVI, EVI, rainfall deviation, growing-degree days (GDD) and rainfall.

This study is not without its limitations. Farmers cannot use the data for management as it was prepared for modelling. It can meet the demands of the algorithm testing, feature engineering, and manuscript preparation, but implementation requires independent multiple-season farm data, validated satellite data time series, district-wide agronomic data (such as fertiliser use) and crop cuts. We must cross-validate on other farms, include sequence prediction of weather and vegetation index, measure the development district-wide and train several models for crop families with different yield. We also should add uncertainty, explainable recommendations and retraining to the app to adapt to changes in climate and agronomy. The proposed approach is technically acceptable for predictive systems, but will not be accepted in the field without rigorous field-scale validation, data handling and iterative agronomic re-calibration. Validation should be evaluated as a modelling system, rather than a "golden" system. This is the next-generation crop yield intelligence under production, based on region, season and management.

REFERENCES

1) M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin, and N. Khan, “A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction,” IEEE Access, vol. 9, pp. 63406–63439, 021, doi: 10.1109/ACCESS.2021.3075159.

2) Paudel, H. Boogaard, A. de Wit, S. Janssen, S. Osinga, C. Pylianidis, and I. N. Athanasiadis, “Machine learning for large- scale crop yield forecasting,” Agricultural Systems, vol. 187, Art. no. 103016, Feb. 2021, doi: 10.1016/j.agsy.2020.103016.

3) Sagan, M. Maimaitijiang, S. Bhadra, M. Maimaitiyiming, D. R. Brown, P. Sidike, and F. B. Fritschi, “Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 174, pp. 265–281, Apr. 2021, doi: 10.1016/j.isprsjprs.2021.02.008.

4) J. Shook, T. Gangopadhyay, L. Wu, B. Ganapathysubramanian, S. Sarkar, and A. K. Singh, “Crop yield prediction integrating genotype and weather variables using deep learning,” PLOS ONE, vol. 16, no. 6, Art. no. e0252402, Jun. 2021, doi: 10.1371/journal.pone.0252402.

5) P. Hara, M. Piekutowska, and G. Niedbała, “Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data,” Land, vol. 10, no. 6, Art. no. 609, Jun. 2021, doi: 10.3390/land10060609.

6) Y. Ma, Z. Zhang, Y. Kang, and M. Özdoğan, "Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach," Remote Sensing of Environment, vol. 259, Art. no. 112408, Jun. 2021, doi: 10.1016/j.rse.2021.112408.

7) M. Engen, E. Sandø, B. L. O. Sjølander, S. Arenberg, R. Gupta, and M. Goodwin, "Farm-Scale Crop Yield Prediction from Multi-Temporal Data Using Deep Hybrid Neural Networks," Agronomy, vol. 11, no. 12, Art. no. 2576, Dec. 2021, doi: 10.3390/agronomy11122576.

8) Nazir, S. Ullah, Z. A. Saqib, A. Abbas, A. Ali, M. S. Iqbal, K. Hussain, M. Shakir, S. M. Shah, and M. U. Butt, "Estimation and Forecasting of Rice Yield Using Phenology-Based Algorithm and Linear Regression Model on Sentinel-II Satellite Data," Agriculture, vol. 11, no. 10, Art. no. 1026, Oct. 2021, doi: 10.3390/agriculture11101026.

9) P. Muruganantham, S. Wibowo, S. Grandhi, N. H. Samrat, and N. Islam, "A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing," Remote Sensing, vol. 14, no. 9, Art. no. 1990, Apr. 2022, doi: 10.3390/rs14091990.

10) Oikonomidis, C. Catal, and A. Kassahun, "Hybrid Deep Learning-based Models for Crop Yield Prediction," Applied Artificial Intelligence, vol. 36, no. 1, pp. 1–18, 2022, doi: 10.1080/08839514.2022.2031823.

11) Oikonomidis, C. Catal, and A. Kassahun, "Deep learning for crop yield prediction: a systematic literature review," New Zealand Journal of Crop and Horticultural Science, vol. 51, no. 1, pp. 1–26, 2023, doi: 10.1080/01140671.2022.2032213.

12) Nyéki and M. Neményi, "Crop Yield Prediction in Precision Agriculture," Agronomy, vol. 12, no. 10, Art. no. 2460, Oct. 2022, doi: 10.3390/agronomy12102460.

13) J. Wang, H. Si, Z. Gao, and L. Shi, "Winter Wheat Yield Prediction Using an LSTM Model from MODIS LAI Products," Agriculture, vol. 12, no. 10, Art. no. 1707, Oct. 2022, doi: 10.3390/agriculture12101707.

14) Tesfaye, B. G. Awoke, T. S. Sida, and D. E. Osgood, "Enhancing Smallholder Wheat Yield Prediction through Sensor Fusion and Phenology with Machine Learning and Deep Learning Methods," Agriculture, vol. 12, no. 9, Art. no. 1352, Sept. 2022, doi: 10.3390/agriculture12091352.

15) Y. Shen, B. Mercatoris, Z. Cao, P. Kwan, L. Guo, H. Yao, and Q. Cheng, "Improving Wheat Yield Prediction Accuracy Using LSTM-RF Framework Based on UAV Thermal Infrared and Multispectral Imagery," Agriculture, vol. 12, no. 6, Art. no. 892, Jun. 2022, doi: 10.3390/agriculture12060892.

16) S. Fei, M. A. Hassan, Y. Xiao, X. Su, Z. Chen, Q. Cheng, F. Duan, R. Chen, and Y. Ma, "UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat," Precision Agriculture, vol. 24, no. 1, pp. 187–212, Feb. 2023, doi: 10.1007/s11119-022-09938-8.

17) K. Jhajharia, P. Mathur, S. Jain, and S. Nijhawan, "Crop Yield Prediction using Machine Learning and Deep Learning Techniques," Procedia Computer Science, vol. 218, pp. 406–417, 2023, doi: 10.1016/j.procs.2023.01.023.

18) S. Iniyan, V. Akhil Varma, and C. Teja Naidu, "Crop yield prediction using machine learning techniques," Advances in Engineering Software, vol. 175, Art. no. 103326, Jan. 2023, doi: 10.1016/j.advengsoft.2022.103326.

19) M. Morales and F. J. Villalobos, "Using machine learning for crop yield prediction in the past or the future," Frontiers in Plant

20) Science, vol. 14, Art. no. 1128388,2023, doi: 10.3389/fpls.2023.1128388.

21) U. Shafi, R. Mumtaz, Z. Anwar, M. M. Ajmal, M. A. Khan, Z. Mahmood, M. Qamar, and H. M. Jhanzab, "Tackling Food Insecurity Using Remote Sensing and Machine Learning-Based Crop Yield Prediction," IEEE Access, vol. 11, pp. 108640–108657, 2023, doi: 10.1109/ACCESS.2023.3321020.

22) Z. Ramzan, H. M. S. Asif, I. Yousuf, and M. Shahbaz, "A Multimodal Data Fusion and Deep Neural Networks Based Technique for Tea Yield Estimation in Pakistan Using Satellite Imagery," IEEE Access, vol. 11, pp. 42578–42594, 2023, doi: 10.1109/ACCESS.2023.3271410.

23) P. Sharma, P. Dadheech, N. Aneja, and S. Aneja, "Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning," IEEE Access, vol. 11, pp. 111255–111264, 2023, doi: 10.1109/ACCESS.2023.3321861.

24) L. Schumacher, E. K. Burchfield, B. Bean, and M. A. Yost, "Leveraging Important Covariate Groups for Corn Yield Prediction," Agriculture, vol. 13, no. 3, Art. no. 618, Mar. 2023, doi: 10.3390/agriculture13030618.

25) S. R. Gopi and M. Karthikeyan, "Effectiveness of Crop Recommendation and Yield Prediction using Hybrid Moth Flame Optimization with Machine Learning," Engineering, Technology & Applied Science Research, vol. 13, no. 4, pp. 11360–11365,

Aug. 2023, doi: 10.48084/etasr.6092.

26) Raju, D. V. Ashoka, and B. V. Ajay Prakash, "CropCast: Harvesting the future with interfused machine learning and advanced stacking ensemble for precise crop prediction," Kuwait Journal of Science, vol. 51, no. 1, Art. no. 100160, Jan. 2024, doi: 10.1016/j.kjs.2023.11.009.

27) M. Rajakumaran, G. Arulselvan, S. Subashree, and R. Sindhuja, "Crop yield prediction using multi-attribute weighted tree-based support vector machine," Measurement: Sensors, vol. 31, Art. no. 101002, Feb. 2024, doi: 10.1016/j.measen.2023.101002.

28) L. K. Subramaniam and R. Marimuthu, "Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops," e-Prime—Advances in Electrical Engineering, Electronics and Energy, vol. 8, Art. no. 100611, 2024, doi: 10.1016/j.prime.2024.100611.

29) U. V. Nikhil, A. M. Pandiyan, S. P. Raja, and Z. Stamenković, "Machine Learning-Based Crop Yield Prediction in South India: Performance Analysis of Various Models," Computers, vol. 13, no. 6, Art. no. 137, Jun. 2024, doi: 10.3390/computers13060137.

30) S. Yang, L. Li, S. Fei, M. Yang, Z. Tao, Y. Meng, et al., "Wheat Yield Prediction Using Machine Learning Method Based on UAV Remote Sensing Data," Drones, vol. 8, no. 7, Art. no. 284, Jul. 2024, doi: 10.3390/drones8070284.

31) S. Jeong, J. Ko, J.-O. Ban, T. Shin, and J.-M. Yeom, "Deep learning-enhanced remote sensing-integrated crop modeling for rice yield prediction," Ecological Informatics, vol. 84, Art. no. 102886,Dec.2024,doi:10.1016/j.ecoinf.2024.102886.