Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 5s (2025)

Integrating Argonomic Method and Yield Crop Modelling with ML Baseline

Dr.V.Sujatha [1], R SHALINI [2], Dr.A.M.Arun Mohan [3],Dr.B.Komala Durga [4], Adarsh Rai [5],Dr Subba Rao Polamuri [6]

[1] Department of Computer Science and Engineering, Vignan's Nirula Institute of Technology and Science for Women, Guntur, India. sujiekkurthi@gmail.com

[2] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fileds, Vaddeswaram, India. spshalinikaran@kluniversity.in

[3] Department of Civil Engineering, Sethu institute of Technology, Pulloor,Kariapatti, India. arunmohan@sethu.ac.in

[4] Department of Mathematics, Virudhunagar Hindu Nadars' Senthikumara Nadar College, Virudhunagar, India. komaladurga.b@vhnsnc.edu.in

[5] Nitte (Deemed to be University), NMAM Institute of Technology, Nitte, Karkala, India. adarsh.rai@nitte.edu.in

[6] Department of Computer Science and Engineering, Aditya University, Surampalem, India. psr.subbu546@gmail.com

Article History:

Received: *11-10-2024*

Revised: *27-11-2024*

Accepted: *09-12-2024*

Abstract:
Predicting agricultural yields using machine learning has been the subject of several research, most of which have concentrated on individual cases. It is possible that their data and techniques won't work with other crops or in different places. Contrarily, machine learning is not used by operating large-scale structures like MARS Crops Harvest Foretelling Systems (MCYFS). When there is a flood of newly disclosed data, ML becomes a potentially useful tool. In order to construct a machine learning foundation for huge-scale crop yield prediction, researchers integrated agronomic concepts of crop modelling with machine learning. A methodology that focusses on accuracy, modularity, and reusability forms the basis. Researchers made sure everything was accurate by using machine learning without leaking any information and by creating characteristics or predictors that could be explained in connection to the development and growth of crops. The characteristics were developed by researchers using information from the MCYFS database, which included meteorological, remote sensing, soil, and crop simulation results.

The researcher focused on a reusable and modular procedure that can accommodate various crops and nations with little configuration adjustments. Using standard input data, the method may be utilised to conduct repeated tests and produce repeatable outcomes. The outcomes act as a springboard for more improvements. In order to compare the performance, the researchers used a straightforward strategy that required no prediction skills and either projected the training set's average for 3 countries like India (IND), Spain (SP) and Ireland (IL) and 5 crops like springs barley, wheat, sunflower, potatoes and sugar beetroot). The researcher forecasted yield at the regional level in the case studies. Additionally, the researchers compared the projections with previous MCYFS forecasts and aggregated them at the national level. Across nations, the normalised RMSE (NRMSE) for the beginning season forecasts was similar. Wheat had an NRMSE of 7.88 for wheat (IND), whereas sweet beetroot had an NRMSE of 8.22 (SP). NRMSEs for potatoes, sugar beetroot and wheat, on the other hand, were double those of MCYFS. At the conclusion of the season, NRMSEs were still similar to MCYFS. Adding more information foundations, creating high analytical characteristics & testing various ML algorithms may all help to enhance the baseline. Large-scale agricultural harvest prediction using ML will be encouraged by the baseline.

Keywords: ML, Crop, MCYFS, Wheat, Sugar Beetroot

Abstract:

1. Introduction:

Predicting agricultural production has traditionally included using field surveys, models of crop growth, remote sensing, statistical models, or a mix of these methods. Each of these approaches deals with a little different facet of crop production prediction on its own. Capturing the ground reality is the goal of field surveys. Models for crop development and growth that take into account the interplay between plants, their environments, and their managers are known as crop growth models. Using satellite data, remote sensing systems may assess the present condition of crops and predict their ultimate output. In order to develop linear correlations among the predictors and the yield of crops, arithmetical replicas make utilize of meteorological values and the outcomes of the 3 preceding approaches as interpreters. In order to construct yield forecasting models, recent research has creatively integrated several approaches.

Machine learning offers a viable way to enhance crop production forecasts by using an empirical modelling or data-driven method to identify meaningful patterns and correlations in input data. A function that connects characteristics to tags, such crop yield, is approximated by machine learning techniques. The results of alternative techniques may be used as features by machine learning algorithms, much as in statistical models. Moreover, there are a number of clear advantages to machine learning algorithms, including the ability to models non-linear relationships between various data sources, the fact that they perform better overall with more training data, and the ability to become resilient to noisy data through the use of regularisation techniques that reduce variance and generalisation error. In order to provide accurate agricultural production forecasts, machine learning may integrate the advantages of existing techniques, such as remote sensing and crop growth models, with data-driven modelling.

To better recognize the utility of different information sources, interpreters and ML techniques of various crops across geographical & progressive contexts, this study aims to talk about the requirement for modular & recyclable processes. With reusable processes, scientists might use standard input data to conduct repeated tests with consistent outcomes, like early season & season end forecasts, across many yields & nations. Through the incorporation of fresh data sources, enhanced features, and various optimisations, the models have the potential to be enhanced for certain crops and geographical areas.

Machine learning isn't used by operational systems researchers, however. Weather data, field examination findings, yield development design results, remote detector & harvest statistics are all used to create statistical models. A possible approach is machine learning, particularly when a lot of information is being gathered and made available to the public. Machine learning would be encouraged to be used in large-scale agricultural yield forecasting via a returnable & extendable approach with respect to inputs comparable to MCYFS.

Researcher provide a ML foundation for large-scale predictions of agricultural yields at the beginning and conclusion of the season. A generic machine learning methodology with an emphasis on (i) accuracy, (ii) modularity, and (iii) reusability serves as the foundation. The primary goals of our approach are using ML in a way that prevents data leakage from the test set and developing attributes that can explain the development and growth of crops according to agronomic concepts from crop modelling. Second, a modular architecture makes it easy to include new data sources, build better features, and test out various machine learning approaches, all of which may enhance or expand the process. Thirdly, reusability takes into account the fact that the process may be easily adapted to other crops and regions by making little adjustments to the setup. Additional optimisations might be based on the given findings.

The ML was tried on 5 crops like spring barley, sunflower, wheat, sugar beetroot & potatoes from 3 nations such as IND, SP and IL. To guess the crop harvest at the NUTS2 and NUTS3 at the beginning and end of the season, researcher did tests. A simple method that can't make predictions, that is called as the "null" method, was used to compare the area forecasts. Either a straight-line trend in yield or the mean of the training dataset was what the null technique estimated. To make things even better, researcher compared the data to previous MCYFS expectations at the national (NUTS0) level.

The rest of the article is structured like this: Part 2 lays out the process; Part 3 shows the outcomes; Part 4 talks about the outcomes and potential future study directions; and Section 5 gives a brief summary of our findings.

2. Methodology:

The researchers used MCYFS data to create a machine learning methodology for agricultural yield prediction. Using A2 or A3 agricultural yield predictions for 5 crops and 3 nations, the researcher assessed the procedure. Researchers experimented with utilising the anticipated yield trend from prior years to forecast early season and season end crop yields for each crop & nation. The "null" technique, which is a straightforward approach that requires no prediction expertise, was used to assess the regional predictions for each experiment. Additionally, the predictions were aggregated to the national S0 levels and compared with previous MCYFS forecasts.

The researcher built the process using precision, modularity, and reusability as its guiding principles. Two components make up the whole process. Feature design and pre-processing make up the first section. ML is a component of the second section.

2.1. Architecture of work processes:

2.1.1 Precision - ML without Data Outflow:

Crop yield prediction was done by the researcher using supervised learning, most especially supervised regression. Training examples that have both features and labels, like yield statistics, are essential for supervised learning in order to teach a purpose which connects attributes with tags. To create training and test sets, the researchers divided the whole dataset. The researcher expanded the test dataset to comprise the last several years used for each location in order to use the yield trend (figure 1). This restriction was necessary because include earlier years into the assessment sample would lead to knowledge leakage, and subsequent years would incorporate yields trend estimates from earlier years.

The researcher may have adopted random splits in place of the yield trend analysis. For comparing the predictions with MCYFS, however, the researchers need identical test years for every location. Thus, using n as defined by the test percent, the researcher added to the test set every nth year. 29% of the data was set aside for testing and 71% for training in both scenarios.

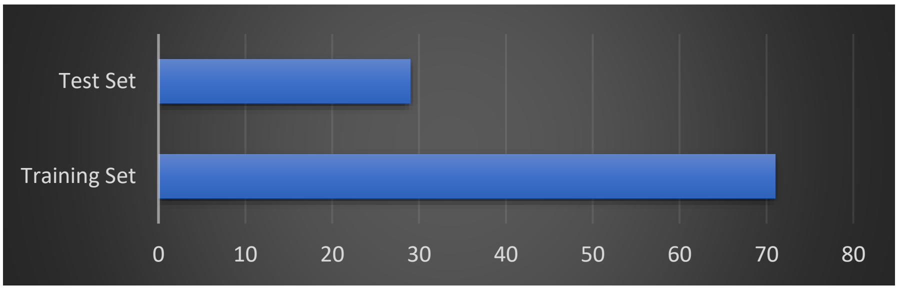

Figure 1 illustrates the division of training dataset & test datasets.

Researcher utilised the test dataset for the last assessment and the training dataset to train & optimise a system. The researchers separated the information into test and training sets before beginning feature design. Only the training set was used by the researcher to compute indicator statistics and infer the crop calendar. In order to prevent data loss during training and feature selection, researchers-built pipelines that included scaling features, selecting features, and training. Each optimisation and training step was guaranteed to utilise only training data by the pipelines. In essence, the training set taught the model how to scale features, how many features to use, and how to weight features. The hyperparameters were fine-tuned with the help of the training set alone, according to the researchers. The pipeline was executed for every iteration of 6-fold cross-validations while optimising the hyperparameters. For this reason, researcher used the training folds to execute the pipeline steps and the test folds to assess the trained model.

2.2. Modularity:

The researcher concentrated on keeping the baseline comparatively simple to expand and enhance in order to promote modularity. Researcher reduced the interdependence between the workflow's subsequent steps. To enable the pointers used for attributes plan to alter without impacting the process, the researcher selected extensible data structures. Enhancing current ones with fresh information served as the aim. Extreme circumstances characteristics, for instance, count days with values between plus or minus one and plus or minus two SD from the mean value. The method is general and reusable due to the usage of indicator mean and SD. However, such information might be useful to physically build high precise and analytical features if crop-specific thresholds for various pointers are known.

Figure 2 shows the setting choices that the researchers used to regulate the flow of data during the different tests. As an example, geographical data on the centres of regions was optionally provided but not included by default. The procedure remained same; but, by modifying the configuration settings and executing the workflow, other experiments could be done. More so, machine learning could be integrated into the workflow without requiring preprocessing or feature design if the produced features were stored in a file and then uploaded to use in ML. In a similar vein, MCYFS might be used to compare machine learning algorithm predictions that were stored to a file.

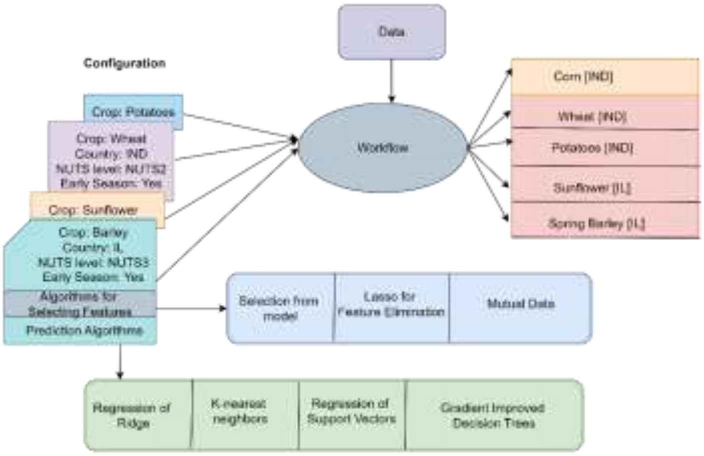

Figure 2 Schematic diagram for test configurations.

It's possible to try out different algorithms because researcher designed the feature selection and forecast algorithms in a way that can be expanded and modified. Given the number of attributes to choose, feature selection methods could be added. Set some hyperparameters to their fixed values & tell the computer the variables for various hyperparameters for optimise. This way, forecast algorithms could be added. Researcher made listings that could be grown or shrunk to show a range of values for hyperparameters.

2.3. Reusability:

Researcher created a procedure that can be applied to many crops and nations. To reduce the quantity of input needed to execute the process, data homogenisation is used to standardise the filenames, file formats, and data fields. For many case studies, identical feature design is used. The procedure was made feasible by data homogenisation and configuration choices for crop name, country, and NUTS level. The researchers consulted a wide range of sources, including the results of the WOFOST crop growth model, data from weather stations, soil samples, regional centres, modelled crop area fractions, and yield figures from other nations. The researchers trained their model using 71% of the data, whereas they tested it using 29%. Since it was not apparent if region centroids added information not found in WOFOST results and weather observations, they were not utilised by default.

Researcher used 14 case studies and conducted five tests for each to confirm the machine learning workflow's precision, flexibility, and reusability. In order to confirm that the characteristics are explainable, researchers first tallied the frequency of certain attributes for every crop across various nations and algorithms. A thorough examination of feature significance was postponed for further study. Second, for each crop and nation, researchers conducted five tests with choices for utilising production trends (Yes or No) and early seasons forecast (Yes or No) in order to confirm the workflow's flexibility. Researchers utilised information from the current season for a maximum of thirty days after planting to anticipate the early season. Up to the conclusion of the harvest window, researcher utilised current season data to make end-of-season predictions. Third, researchers conducted five trials for fourteen case studies consisting of potatoes, sugar beetroot, sunflower, wheat and spring barley in order to confirm reusability. Wheat was used to evaluate the workflow's optional components. Estimates were made at NUTS2 & NUTS3 for different nations. Overall, researcher used five crops, two NUTS levels, and several nations to evaluate the procedure.

Ridge Regression, K-nearest Neighbours Regression, SVM Regression & Gradient Boosted Decision Trees Analysis were the four machine learning algorithms that are tested for their ability to forecast crop production. Different algorithms' approaches to learning feature-label associations are reflected in these strategies. ML algorithms' predictions are contrasted with a basic, skill-free technique. Without yield trend, the null technique was just the Zero R algorithm, which takes the training set average into account while making predictions. The linear harvest pattern derived from a five-year period was predicted by the null approach if harvest pattern is employed. The reliability, validity, and accuracy of each algorithm were assessed by calculating their MAE, MPE, RMSE, and coefficient of determination. The normalised versions of MAE & RMSE were used to compare the two. In order to get the normalised errors, researcher divided the mean error by the mean yield of the information sample.

2.4. Compared to the predictions of MCYFS:

To total the crop harvest forecasts from NUTS0 to NUTS3, researcher had to modify the training and test divides. The test dataset needed to comprise a similar set of years for every area. Certain locations and test years did not appear in forecasts when researcher set the test years to be the same. Two methods are used to fill in the missing predictions. First, the mean for the remaining years is used to fill in the missing number if the area had projections for previous test years. Second, the area percentages of other siblings' areas and disregarded the region are modified if it had no forecasts at all.

2.5. Execution:

For data preparation and feature creation, Apache Spark dataframes is used. To apply machine learning, the scikit-learn python tool is employed. Google Colaboratory is used to build and test the process, and finally the Google Dataproc cluster & Microsoft Azure Databricks are utilized to conduct the various tests.

3. Results:

For each crop, researchers examined feature selection frequencies across several nations and techniques to confirm the explainability of characteristics. To illustrate modularity and reusability, researchers conducted 5 trials for each of the 15 crop & country groupings like spring barley, sunflower, wheat, sugar beetroot, and potatoes with the choices to utilise yield pattern (Yes/No) & to forecast early in the season (Yes/No). Future projections at NUTS2 and NUTS3. The findings were compared to previous MCYFS projections and totalled at the national level. For several case studies, the researchers offer the normalised RMSE in this section.

3.1. The frequency of feature selection:

Water-holding capacity of the soil was always chosen, according to feature selection criteria for potatoes. Likewise, every characteristic for the pre-planting time was chosen a lot. The average and extreme temperatures and precipitation were significant during the planting window. For the vegetative phase, the most often chosen characteristics were the average temperature, water-limited yielding biomass, leaf surface index, and percent of absorbed photosynthesis-active radiation (FAPAR). The blossoming phase depended on precipitation and extremes in maximum temperature. Water-restricted harvest biomass, total water usage & yield storing were all crucial FAPAR & WOFOST metrics during the yield formation phase. Lastly, the harvest window was influenced by the average and exceptional precipitation. In general, the variables influencing the development of crops during these times align with feature selection frequencies. Extremes in precipitation during the planting and harvest windows, as well as temperature during the blooming period, are known to affect crop production.

3.2. Trends in yield as opposed to no trend:

Researchers contrasted the null method's production pattern (Yes or No) end-of-season forecasts with those using the Gradients Boosted Decision Diagrams (GBDT) approach (figure 3). For the most part, GBDT outperformed competing algorithms, therefore researcher went with it. Machine learning outperformed the null technique in most cases, with the exception of sugar beets and MAPE for potatoes, which showed a "Yield Trend" rather than a "No Yield Trend". This "Yield Trend" vs. "No Yield Trend" comparison is flawed due to the fact that the two sets of data used for training and testing are different. However, there was little difference between the two sets of error data, suggesting that machine learning may be used with or without productivity trends. The test set for the yield trend includes the last years that were available. For this reason, forecasts based on a yield pattern would be practical. For years that are absent in the forecast, the "No Yield Trend" method may be helpful.

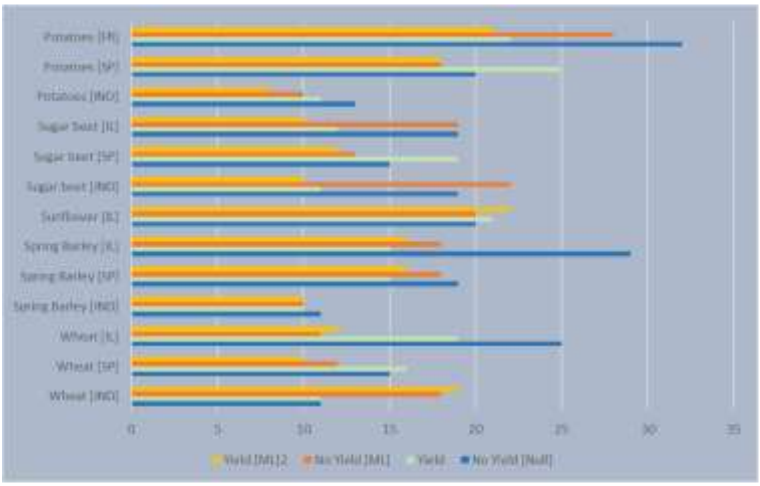

Figure 3 depicts comparison of yield pattern vs no yield pattern.

3.3. Comparison between season-ending and beginning estimates:

Using yield trend to make early season estimates (figure 4) showed that the baseline method might make better early season forecasts than the null technique. Researcher chose to compare GBDT to other algorithms because, in most situations, it worked better than the others. In every case except for MAPE for potatoes (FR), the normalised RMSE and MAPE outcomes for ML was lower compared to those for the null technique. The null method used a linear 5-year pattern to guess the return. Predictions for the early season were developed 30 days after gardens were planted. When the harvest window closed, statements were made about the season end. Estimation for the beginning & end of the season both used data from the earth, five previous years' yields, and information about the current season up to the forecast dekad. The machine learning standard error numbers got a little better across the duration of the season, except for Spring Barley.

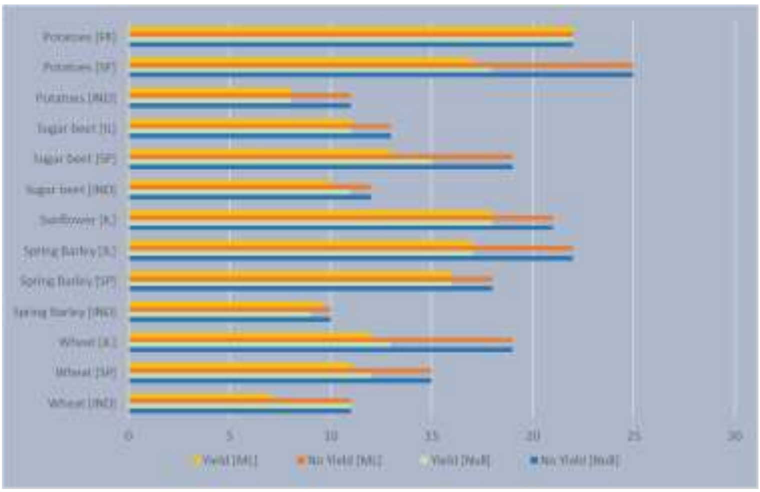

Figure 4 illustrates Early season estimation by five year production pattern.

3.4. Evaluation with MCYFS projections:

The ML initial findings were added up to NUTS0, and they were compared to past MCYFS estimates. Researcher contrasted the estimates for ML algorithms utilising the yield trend because the MCYFS method searches for patterns. Researcher compared the results with estimates from the most efficient machine learning algorithm. The method that was chosen for each case study was different. Looking at the early year, compared MCYFS predictions from the most recent dekad with machine learning predictions for thirty days after planting (figure 5).

Figure 5 demonstrates NUTSO forecasts using yield trend in comparison to MCYFS.

Additionally, researcher looked at the last MCYFS estimate for the year and compared it to estimates made by ML at the close of the harvesting window (figure 6). As the season started, the machine learning average did about the same as MCYFS. All five crops in IND, sugar beets, potatoes, and barley in SP, and wheat, spring barley, and sunflower in IL were predicted to be almost the same. As an example, the Normalised RMSE for wheat (IND) was 7.88 (6.33 for MCYFS), sugar beet (SP) was 8.22 (8.80 for MCYFS), and sunflower (IL) was 10.64 (10.92 for MCYFS). The Normalised RMSE for wheat (SP) was 16.39 (6.22 MCYFS), and for sugar beetroot (IL) was 14.33 (MCYFS 7.43), indicating significantly worse predictions compared to SP and IL, respectively. Figures 4 and 5 show that as the season went on, MCYFS projections became substantially better, whereas machine learning predictions didn't get much better. Normalised RMSE in wheat (IND) (MCYFS 5.49), for example, was 3.06, indicating that IND predictions were similar to MCYFS; however, SP and IL predictions were inferior. Soil data, weather observations, WOFOST outputs, and remote sensing indicators were the same data sources utilised by the baseline all season long. In contrast, MCYFS updates its projections by consulting additional sources of information, such as agricultural periodicals and news articles. Furthermore, MCYFS analysts play a crucial role in determining which underlying feature data best explains crop development & harvests, and in selecting the right statistical system to provide accurate yield predictions.

Figure 6 illustrates the estimation of season end using a 5-year production pattern.

4. Discussion:

Preciseness, flexibility, and reusability were the three main ideas that guided the creation of the machine learning standard. Before researcher used machine learning, we made sure that the features we designed could be explained and that there was no information leaking. Feature sets made from numbers from past years, like the yield pattern, are used if analysing time series data, like crop yield. There needs to be extra care taken to keep information from earlier years from getting out when it's included in features. To make sure that data from the test dataset is not utilised during training, the baseline has a time-dependent training & testing divide and a k-fold slide authentication. To let the process change over time and to try out different setups, we put a lot of emphasis on flexibility. For each case study, the process allows for small changes that make the baseline longer and better. Finally, we made sure that the process could be used again and again for different things and places. There will be fewer centralised and similar software solutions because of the focus on flexibility and reusability. This will increase the reuse of models and software.

The feature design process, which continues by picking attributes later in the workflow, is one of the baseline's major innovations. Using agronomic concepts from crop modelling, researcher created characteristics. Researchers determined which indicators had an impact on crops at various times of the crop cycle. In order to cater for harsh situations, researcher additionally implemented features. The method was general and reusable since features for severe situations depended on the mean & standard deviations of indicators. Researcher investigated the range of thresholds for severe situations by generating a huge quantity of attributes, & researcher used attribute choice to determine the proper thresholds. In a similar vein, choose the best predicted ones by hand rather than relying on experts. In order to identify the characteristics that account for yield variability for every crop and nation, a data-driven method is utilized.

In order to forecast agricultural production, researcher used supervised machine learning, that is very data-dependent, and ran the baseline. When the training set is an accurate representation of the whole dataset and the training labels are trustworthy, supervised learning algorithms perform very well as predictors. Researcher opted to forecast agricultural production on a sub-country scale & pooled information from many locations to guarantee a substantial set. National-level MCYFS projections are based on data on agricultural yields. Statistics on yields at the subnational level are less often updated and might differ across nations and types of crops.

Data from prior years is duplicated in some locations, while missing data is present in others. Thus, the trade-off between data amount and data quality is shown by regional agricultural production forecast. However, particularly in early season, the combined NUTS0 ML projections were encouraging. The baseline output was similar to MCYFS for IND (all 4 crops), SP (spring barley, potatoes and sugar beetroot), & IL (wheat, sunflowers and spring barley) (see figure 5). In terms of approach, MCYFS trains models for the next year using data from every prior year. In contrasting the outcomes of the baseline & the MCYFS projections, such variations in data and methodology should be taken into account. To further understand how machine learning might enhance agricultural production estimates, future studies should analyse the effects of various features, algorithms, hyperparameters, and regularisation approaches, as well as ways to fix data quality. When dealing with specific crops or nations, it may be more prudent to use sub-national level data for agricultural yield forecast rather than national level statistics. There is the possibility that regional analyses and the combined national yield projections will be more precise. These studies would build on top of the machine learning norm.

The baseline might be used right away to crops & nations that are covered by MCYFS because it is based on statistics from MCYFS. Researcher can also use the baseline for situations where similar crop development and yield markers can be obtained from other crop modelling models. These could be dry mass harvest biomass, leaf surface, or growing phase. If information about the balance of the countries comes in a manner alike to MCYFS, the machine learning standard will be useful.

In terms of fit-for-purpose optimisations and basic design principles, the baseline offers plenty of space for development. According to our experience, there are at least five approaches to enhance the baseline. First, the standard of training data may be enhanced by the identification of outliers & duplicate data. Second, to create a more optimised machine learning model, the effects of various features, algorithms, hyperparameters, and regularisation techniques might be examined. Third, by using the proper data homogenisation and preprocessing, additional data sources might be incorporated. Feature design is another point to take into account. Certain data sources may be utilised as features right away, while others need careful feature design. Fourth, certain more information may improve the accuracy of feature design. Researcher use WOFOST results to derive the national agricultural calendar in the baseline. When the nation encompasses many agro-ecological zones, a crop calendar might be created for each area. The crop calendar might be defined by remote sensing, phenological databases, or more precise sowing and harvest dates. Likewise, severe circumstances might be defined by crop-specific criteria. Fifth, more sophisticated features that record changes in patterns of cropping and include climate or soil data from prior years might be created.

Some technological restrictions are also present in the machine learning baseline. To begin, there is no one-size-fits-all approach to data preparation in the baseline. In order to meet the standards of the baseline, data from specific nations and crops may need substantial preprocessing. In addition, extremely massive data analysis does not use the baseline. Researchers used scikit-learning for picking features and machine learning, however Spark data frames were utilised for dispersed preprocessing and feature creation. While optimising hyperparameters or running several algorithms, Scikit-learn does not divide up data and calculations. Feature selection was the primary motivation for using scikit-learn as opposed to the Spark machine learning package. The necessary features may be added to Spark MLlib in a future version. Whatever the case may be, moving the workflow's machine learning operations to a distributed environment might be the subject of future studies.

5. Conclusions:

Researcher created a machine learning methodology for crop production prediction that is reusable and modular, and researcher tested it using case studies. Overall, researchers discovered that crop production can be predicted at the sub-national level utilising explainable characteristics created using crop modelling concepts. In the majority of instances, the ML baseline outperformed MCYFS for early season forecasts. As the season went on, there was potential for improvement. Machine learning-based sub-national yield prediction is a promising future strategy for crops and nations with reliable regional data. In addition to fixing problems with data quality, there are three primary approaches to enhance the baseline: incorporating more information sources, creating high analytical characteristics & testing various algorithms. Investigating ML's potential for large-scale agricultural production forecasting begins with the machine learning baseline.

Reference:

[1] Bregaglio, Simone, et al. "Improving crop yield prediction accuracy by embedding phenological heterogeneity into model parameter sets." *Agricultural Systems* 209 (2023): 103666.

[2] Feng, Puyu, et al. "Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia." *Agricultural and Forest Meteorology* 275 (2019): 100-113.

[3] A. Betts, Richard. "Integrated approaches to climate–crop modelling: needs and challenges." *Philosophical Transactions of the Royal Society B: Biological Sciences* 360.1463 (2005): 2049-2065.

[4] Mirshekarnezhad, Babak. "Integrating agrotechnical practices to optimize maize yield potentials in a regional variable climate: DSSAT and Python tools." *Cereal Research Communications* 52.1 (2024): 301-312.

[5] Shook, Johnathon, et al. "Crop yield prediction integrating genotype and weather variables using deep learning." *Plos one* 16.6 (2021): e0252402.

[6] Kadiyala, M. D. M., et al. "An integrated crop model and GIS decision support system for assisting agronomic decision making under climate change." *Science of the Total Environment* 521 (2015): 123-134.

[7] von Bloh, Malte, David Lobell, and Senthold Asseng. "Knowledge informed hybrid machine learning in agricultural yield prediction." *Computers and Electronics in Agriculture* 227 (2024): 109606.

[8] Reynolds, Matthew, et al. "Role of modelling in international crop research: overview and some case studies." *Agronomy* 8.12 (2018): 291.

[9] Shahhosseini, Mohsen, Guiping Hu, and Sotirios V. Archontoulis. "Forecasting corn yield with machine learning ensembles." *Frontiers in Plant Science* 11 (2020): 1120.

[10] Torgbor, Benjamin Adjah, et al. "Integrating remote sensing and weather variables for mango yield prediction using a machine learning approach." *Remote Sensing* 15.12 (2023): 3075.

[11] Zhang, Liangliang, et al. "Planning maize hybrids adaptation to future climate change by integrating crop modelling with machine learning." *Environmental Research Letters* 16.12 (2021): 124043.

[12] Kumar, K. Nanda, Adarsh Vijayan Pillai, and MK Badri Narayanan. "WITHDRAWN: Smart agriculture using IoT." (2021).

[13] Crane-Droesch, Andrew. "Machine learning methods for crop yield prediction and climate change impact assessment in agriculture." *Environmental Research Letters* 13.11 (2018): 114003.

[14] Basso, Bruno, and Lin Liu. "Seasonal crop yield forecast: Methods, applications, and accuracies." *advances in agronomy* 154 (2019): 201-255.

[15] Rajulu, G. Govinda, M. Jamuna Rani, D. Deepa, Udit Mamodiya, Radhika G. Deshmukh, and T. Rajasanthosh Kumar. "Cloud-Computed Solar Tracking System." *Computer Communication, Networking and IoT* 2 (2021): 75.

[16] Morales, Alejandro, and Francisco J. Villalobos. "Using machine learning for crop yield prediction in the past or the future." *Frontiers in Plant Science* 14 (2023): 1128388.

[17] Mewada, Shivlal, et al. "Smart diagnostic expert system for defect in forging process by using machine learning process." Journal of Nanomaterials 2022.1 (2022): 2567194.

[18] Resop, Jonathan P., et al. "Combining explanatory crop models with geospatial data for regional analyses of crop yield using field-scale modeling units." *Computers and electronics in agriculture* 89 (2012): 51-61.

[19] P. Patro, R. Azhagumurugan, R. Sathya, K. Kumar, T. R. Kumar and M. V. S. Babu, "A hybrid approach estimates the real-time health state of a bearing by accelerated degradation tests, Machine learning," 2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2021, pp. 1-9, doi: 10.1109/ICSTCEE54422.2021.9708591.

[20] Mamassi, Achraf, et al. "A comparison of empirical and mechanistic models for wheat yield prediction at field level in Moroccan rainfed areas." *in silico Plants* 6.1 (2024): diad020.

[21] Grassini, Patricio, et al. "Integrating agronomy and machine learning to generate high-resolution global maps of yield potential with local relevance." (2024).