Paper Title: Tea Crop Forecasting: Blending Nature's Wisdom with AI Ingenuity

Paper Link : [Plants | Free Full-Text | A Hybrid Approach to Tea Crop Yield Prediction Using Simulation Models and Machine Learning (mdpi.com)]

Summary: Tea is a highly consumed beverage globally, and accurate and timely tea yield prediction is crucial. Previous studies have used statistical, deep learning, and machine learning techniques, but crop simulation models have not been used yet. This research study aims to compare methods for tea yield prediction using the Food and Agriculture Organization's AquaCrop simulation model and machine learning techniques. The study used data from 2016 to 2019 from the National Tea and High-Value Crop Research Institute in Pakistan. The AquaCrop model was calibrated using weather, soil, crop, and agro-management data. The machine learning regression algorithm performed better in yield prediction using fewer data than the simulation model. This study provides a technique to improve tea yield prediction by combining different data sources using a crop simulation model and machine learning algorithms.

Motivation: Agriculture is a crucial source of national income for many developing countries, with over 50% of the population earning their livelihood from it. Pakistan, with its diverse land and seasons, is well-suited for agriculture, with 96.9% land area and 3.1% water bodies. The agricultural sector contributes significantly to Pakistan's Gross Domestic Product by increasing exports. However, Pakistan's economy relies on balancing imports and exports, and it imports many agricultural products, including tea, from other countries. Tea is a global beverage with a 4.4% annual increase in world tea yield, reaching 5.73 million tons in 2016. The global production of black tea is projected to rise by 2.2% annually over the next ten years, reaching 4.4 million tons in 2027. Green tea production is projected to increase by 7.5% compared to black tea, reaching 3.6 million tons in 2027. Pakistan, the 2nd largest tea importer, imports tea from 21 countries, resulting in a huge import bill. To address this issue and increase tea production, the National Tea Research Institute under the Pakistan Agricultural Research Council (PARC) was established in 1986. The institute has a tea garden, an equipped laboratory, and a tea-processing unit to process tea leaves. Various technology-based methods are utilized for crop yield estimation, including remote sensing, crop simulation models, and statistical methods. These techniques also aid in determining the impact of climate on crop yields. This study aims to improve tea yield prediction in Pakistan by using weather, crop, soil, and agro-management data from 2016 to 2019. The AquaCrop simulation model, which requires fewer parameters, was recently parameterized for table grapes from 2005 to 2006. The study validates the AquaCrop simulation model by comparing its results with ML algorithms. Using multiple data sources, such as weather, crop, soil, and agro-management, can provide a better yield forecast before plucking tea leaves. This approach can improve tea crop production by providing a forecast before plucking leaves. The study also provides a paradigm for CSMs for tea yield prediction in other regions. The research questions addressed include how to calibrate the selected CSM, how ML regression algorithms perform for tea yield prediction, and which technique from both ML and CSM performs better for tea yield prediction. The paper is organized into sections, including relevant studies, data used, results, discussions, and conclusion.

Contribution: Crop yield estimation is crucial for food security and improving management practices to increase crop production. In the past decade, crop yield estimation has been performed using traditional statistical regression methods, machine learning (ML) techniques, and crop models. ML techniques use environmental and genotype data for prediction, while crop growth models (CSMs) represent real-world experiments. These models help farmers, policymakers, and the government maximize sustainability by providing reliable information about crop production.Crop growth models use plant growth processes and run these processes at multiple scales to find yield during plant growth, resulting from variables such as climate, plant density, crop management, and stress factors. Some models have been calibrated for perennial crops, such as the AquaCrop model, which predicts water requirements for various crops. Machine learning (ML) techniques have been used for crop estimation and classification, with studies achieving an RMSE of 11% for corn yield and yield difference between corn hybrids and genotype or environment data. ML techniques have also been used to predict frost hazard for Zhejiang tea trees, with SVM and artificial neural networks achieving 83.8% and 75% accuracy for frost damage prediction, respectively. A spatiotemporal hybrid model was developed using satellite-derived hydro-meteorological variables from 20 stations across Bangladesh between 1981 and 2020, improving tea yield forecasting with the least relative error value of 11%.Deep learning (DL) is an advanced approach used for estimating crop yields, including soybean, corn, and tea. ML and DL techniques use climate, soil, crop, and satellite data to find patterns and estimate yield. Remote sensing data allows crop status monitoring, and climatic and satellite data are used to predict wheat yield. Statistical models have been used to predict the climate effect on tea yield, with correlation analysis showing that tea yield and climatic variables are correlated. Stochastic frontier analysis has been applied to analyze irrigation water use efficiency in tea farms, saving up to 57.81% of water usage. Recent studies have predicted future projections of tea crops using climatic data, with yields expected to increase in China, Vietnam, and India and decrease in Kenya, Sri Lanka, and Indonesia. A simulation model was developed to estimate tea yield using data from tea cultivar TRI 2025 in Sri Lanka. Methodology: Deep learning (DL) and machine learning (ML) techniques are used to estimate crop yields, including soybean, corn, and tea. These techniques use climate, soil, crop, and satellite data to identify patterns and estimate yield. Remote sensing data allows crop status monitoring, while statistical models predict the climate effect on tea yield. Stochastic frontier analysis has been applied to analyze irrigation water use efficiency in tea farms, saving up to 57.81% of water usage.

Methodology: Recent studies predict future projections of tea crops using climatic data, with yields expected to increase in China, Vietnam, and India and decrease in Kenya, Sri Lanka, and Indonesia. A simulation model was developed to estimate tea yield using data from tea cultivar TRI 2025 in Sri Lanka. The FAO AquaCrop model was used to estimate tea yield throughout crop cycles, considering factors like crop development, transpiration, biomass production, and water deficiency. The model uses meteorological parameters like temperature, rainfall, humidity, solar radiation, and wind speed as input data, and an irrigation schedule to maintain water content and overcome rain deficiency.Results And Analysis: The AquaCrop model was evaluated for simulating crop evapotranspiration (ETa), biomass, and dry yield using statistical parameters such as MAE, MSE error, and RMSE. The model's calibration involved adjusting parameters related to CC, ETa, biomass, and yield. The best values for CCo, CCx, CGC, and CGDD were obtained using data from the 2016-2017 growing season. The model also adjusted for ETa, affecting soil evaporation and plant transpiration. The model was calibrated using a calibration dataset, and a comparison of observed and simulated seasonal tea yields was made. The AquaCrop model's yield values align with observed yields, except in 2019 due to inconsistent plucking of tea leaves. Improved data collection strategies can improve results. Simulated yields are significantly correlated with observed yields, with RMSE, MSE, and MAE values.

Conclusion: This research study acquired data from NTHRI and extracted data from four databases, including weather, crop, soil, and agro-management data. We have selected and calibrated the AquaCrop simulation model for tea yield prediction because it requires fewer parameters than the other simulation models. The AquaCrop model was calibrated using data associated with weather parameters, crop, and soil characteristics to estimate tea crop production. We have trained ML models using this data and observed that the XGBoost regressor outperformed all the other models. We concluded that ML techniques perform better than simulation models by comparing the results of both simulation models and ML techniques. We also concluded that simulation models could be best suited for the estimation of the output of those very expensive and time-consuming experiments if enough input parameters are available. For future studies associated with AquaCrop calibration, it is suggested to use all model variables to perform regression analysis. The results can be further improved by recording more detailed information related to crop and field experiments.

Limitation 1: Dependency on Natural Variables Limitation: Tea crop forecasting, when blending nature's wisdom with AI ingenuity, heavily relies on natural variables such as weather conditions, soil quality, and pest infestations. Despite advancements in AI, predicting and modeling these variables accurately remains challenging due to their inherent complexity and variability. Implication: Unpredictable changes in weather patterns or sudden outbreaks of pests can significantly impact the accuracy of forecasts. The blending approach may struggle to account for sudden and unforeseen natural events, leading to potential inaccuracies in crop predictions.

Limitation 2: Cultural and Regional Variability Limitation: Tea cultivation practices vary across different regions, each with its own set of cultural and traditional approaches. AI models may face challenges in capturing and incorporating these nuances into forecasting algorithms. Localized knowledge, often based on generations of experience, may not easily translate into data-driven models.Implication: The blending of nature's wisdom with AI ingenuity may overlook crucial regional or cultural factors, affecting the accuracy of predictions. Localized insights that play a significant role in tea cultivation may not be adequately represented or considered in the forecasting process

Limitation 3: Dynamic Nature of Agriculture Limitation: Agriculture, including tea cultivation, is inherently dynamic and subject to continuous changes influenced by various factors such as market trends, technological advancements, and evolving farming practices. AI models may struggle to adapt rapidly to these dynamic changes, especially when trying to integrate traditional wisdom and contemporary data-driven approaches. Implication: The blending approach may face challenges in keeping pace with the ever-changing landscape of agriculture. If the forecasting model fails to adapt quickly to emerging trends or shifts in farming practices, its predictive accuracy may diminish over time, limiting its effectiveness in providing reliable forecasts.

Synthesis: In the synthesis of Tea Crop Forecasting, marrying nature's wisdom with AI ingenuity creates a harmonious approach to cultivation. This collaboration between tradition and technology not only enhances the efficiency of tea farming but also fosters a sustainable and resilient agricultural ecosystem. As we navigate the future of tea cultivation, the synthesis of nature and AI stands as a testament to the power of innovation in preserving and advancing our agricultural heritage.