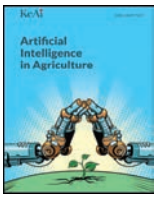
Artificial Intelligence in Agriculture 14 (2024) 99–114

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/



Enhancing crop yield prediction in Senegal using advanced machine learning techniques and synthetic data

Mohammad Amin Razavi [a], A. Pouyan Nejadhashemi [b,*], Babak Majidi [c], Hoda S. Razavi [b], Josué Kpodo [b,d], Rasu Eeswaran [b,e], Ignacio Ciampitti [f], P.V. Vara Prasad [f,g]

[a] School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
[b] Department of Biosystems and Agricultural Engineering, Michigan State University, MI, USA
[c] Department of Computer Engineering, Khatam University, Tehran, Iran
[d] Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, USA
[e] Department of Agronomy, Faculty of Agriculture, University of Jaffna, Kilinochchi, Sri Lanka
[f] Department of Agronomy, Kansas State University, Manhattan, KS, USA
[g] Feed the Future Sustainable Intensification Innovation Lab, Kansas State University, Manhattan, KS, USA

ARTICLE INFO

Article history:
Received 3 August 2024
Received in revised form 5 October 2024
Accepted 8 November 2024
Available online 24 November 2024

Keywords:
Crop yield prediction
Variational auto encoder
Pattern recognition on spatiotemporal and physiographical variables
Synthetic tabular data generation
Ensemble learning

ABSTRACT

In this study, we employ advanced data-driven techniques to investigate the complex relationships between the yields of five major crops and various geographical and spatiotemporal features in Senegal. We analyze how these features influence crop yields by utilizing remotely sensed data. Our methodology incorporates clustering algorithms and correlation matrix analysis to identify significant patterns and dependencies, offering a comprehensive understanding of the factors affecting agricultural productivity in Senegal. To optimize the model's performance and identify the optimal hyperparameters, we implemented a comprehensive grid search across four distinct machine learning regressors: Random Forest, Extreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), and Light Gradient-Boosting Machine (LightGBM). Each regressor offers unique functionalities, enhancing our exploration of potential model configurations. The top-performing models were selected based on evaluating multiple performance metrics, ensuring robust and accurate predictive capabilities. The results demonstrated that XGBoost and CatBoost perform better than the other two. We introduce synthetic crop data generated using a Variational Auto Encoder to address the challenges posed by limited agricultural datasets. By achieving high similarity scores with real-world data, our synthetic samples enhance model robustness, mitigate overfitting, and provide a viable solution for small dataset issues in agriculture. Our approach distinguishes itself by creating a flexible model applicable to various crops together. By integrating five crop datasets and generating high-quality synthetic data, we improve model performance, reduce overfitting, and enhance realism. Our findings provide crucial insights for productivity drivers in key cropping systems, enabling robust recommendations and strengthening the decision-making capabilities of policymakers and farmers in data-scarce regions.

1. Introduction

The United Nations' Sustainable Development Goals focus on enhancing sustainable agriculture, cutting food waste, improving nutrition, and achieving zero hunger (Karthikeyan et al., 2020). With over 7.5 billion people and 700 million facing food insecurity, global food production needs to rise by 50 % by 2050 (Chakraborty and Newton, 2011; Karthikeyan et al., 2020). However, neither arable land nor water is an unlimited commodity. For example, the most recent data on agricultural resources over the past sixty years, show a steady decline in arable land use per capita globally (Ritchie and Roser, 2024) which further attests to the finite and limited availability of the world's arable lands, as observed by the FAO (Kirchmann, 2019). In addition, currently, 80–90 % of global water resources are used for irrigation (Döll et al., 2014; Karthikeyan et al., 2020). Despite expanding agricultural areas, food security in developing nations faces challenges due to inadequate resource management and food and irrigation water pricing policies (Calzadilla et al., 2013; Masson-Delmotte et al.,

* Corresponding author.
E-mail addresses: s.m.amin.razavi@ut.ac.ir (M.A. Razavi), pouyan@msu.edu (A.P. Nejadhashemi), b.majidi@khatam.ac.ir (B. Majidi), razavi@msu.edu (H.S. Razavi), kpodojos@msu.edu (J. Kpodo), rasueesw@msu.edu (R. Eeswaran), ciampitti@ksu.edu (I. Ciampitti), vara@ksu.edu (P.V.V. Prasad).

https://doi.org/10.1016/j.aiia.2024.11.005