value of shrinkage and regularization when many daily weather variables are used. Similar comparisons for rice show that penalized regressions can rival or exceed traditional stepwise regression, though they may still lag behind flexible non-linear models such as neural networks under highly complex climate-yield relationships (Satpathi et al., 2023).

5.2 Machine learning modeling methods

Machine learning (ML) methods such as Random Forest (RF), Support Vector Regression, and boosted trees have become central to crop yield prediction because they capture non-linear responses and interactions between soil, climate, and management variables without strict parametric assumptions. For maize, RF has been shown to outperform multiple linear regression at regional and global scales, reducing RMSE from 14-49% of mean yield with linear models to 6-14% with RF, and better reproducing spatial patterns of yield (Jeong et al., 2016). In the U.S. Midwest, a comparative study using Lasso, Support Vector Regressor, RF, and XGBoost with hundreds of environmental features found that XGBoost was the most accurate and stable algorithm for county-level maize yield prediction (Kang et al., 2020).

In some applications, ML models trained on relatively simple climate inputs also perform strongly. For Irish potato and maize in Rwanda, Random Forest using only rainfall and temperature achieved R^2 values of 0.875 and 0.817, respectively, outperforming polynomial regression and Support Vector Regressor and providing practically useful early-season predictions (Kuradusenge et al., 2023). ML has also been used to model silage maize yields from NDVI time-series; boosted regression trees and RF achieved correlations above 0.87, and were less sensitive to inconsistencies in satellite-derived vegetation profiles than conventional regressions (Aghighi et al., 2018). These studies underline the versatility of ML methods for integrating climate, soil, and remote-sensing predictors in maize yield models.

5.3 Deep learning and ensemble learning methods

Deep learning (DL) extends ML by learning complex, hierarchical representations from large, high-dimensional datasets composed of weather, soil, genotype, and remote sensing inputs. A deep neural network trained on thousands of maize hybrid trials across more than 2,000 locations substantially outperformed Lasso, shallow neural networks, and regression trees, reaching an RMSE close to 11-12% of average yield while also supporting feature selection to reduce input dimensionality with minimal accuracy loss (Khaki and Wang, 2019). However, DL does not always dominate: in a U.S. Midwest maize study, LSTM and CNN architectures did not surpass XGBoost, suggesting that tabular environmental datasets may not always benefit from image- or sequence-oriented deep architectures (Kang et al., 2020).

Ensemble learning combines multiple base learners to improve robustness and accuracy. For corn in the U.S. Corn Belt, CNN-DNN ensembles created via bagging and stacking outperformed ensembles of linear regression, Lasso, RF, XGBoost, and LightGBM, explaining about 77% of spatio-temporal yield variation with an RMSE of 866 kg/ha (Shahhosseini et al., 2021). Hybrid and ensemble DL frameworks that fuse convolutional, recurrent, and fully connected networks have also shown superior performance for crop yield prediction, with CNN-DNN or CNN-RNN-LSTM structures often exceeding single DL or ML models and achieving R^2 values near or above 0.85 in case studies (Oikonomidis et al., 2022). Deep ensemble approaches thus offer a promising route for integrating multi-source soil, climate, and remote-sensing data to achieve robust maize yield prediction under variable environments.

6 Model Training and Evaluation System

6.1 Dataset partitioning and validation strategies

A reasonable partition of the maize yield dataset is the basis for constructing reliable prediction models. In most supervised learning settings, data are divided into training, validation, and test subsets so that model fitting, hyperparameter tuning, and final performance assessment can be clearly separated and avoid information leakage (Bischl et al., 2021). When the number of yearly observations is small, directly reserving an independent test set