reflect operational performance and were used, for instance, in multi-farm machine-learning models and in Ghanaian RF models for maize yield and agronomic efficiency (Filippi et al., 2019; Asamoah et al., 2024).

When rigorously validated, yield prediction models support several agricultural applications. Plot-scale maize models that accurately forecast yield under different fertilizer systems enable assessment of input strategies and refinement of site-specific recommendations before harvest (Meng et al., 2021). Large-area models that integrate climate, soil, and satellite indicators have been used for early-season yield forecasting, outperforming official forecasts and providing actionable information for logistics, market planning, and food-security assessments (Li et al., 2022). Such applications demonstrate how reliable maize yield prediction, grounded in soil-climate interactions, can inform precision fertilization, risk management, and regional policy decisions.

8 Results Analysis and Discussion

8.1 Contribution analysis of soil nutrient variables

Feature-importance and interpretable ML studies highlight that specific soil nutrients can dominate maize yield responses, even in data-rich settings. In a data-intensive farm management trial, Random Forest analysis showed that urea application was consistently the most critical variable for explaining spatial yield variation, with soil phosphorus, pH, clay content, sodium and plant population also among the leading contributors in different seasons (Maseko et al., 2024). This indicates that both applied N and inherent soil fertility properties jointly control yield in high-resolution, within-field prediction. Similar work in precision agriculture, using RF and other models on over 145,000 corn and soybean yield observations, found that soil test P, K, Zn, soil organic matter and cation exchange capacity were key predictors, underscoring the strong explanatory power of nutrient and related soil indicators for yield variation at sub-field scales (Burdett and Wellen, 2022).

Under nutrient-limited conditions, omission trials combined with AutoML provide a more explicit decomposition of nutrient contributions. In 324 nutrient omission plot trials across ten agroecological zones in the Eastern Indo-Gangetic Plains, stack-ensemble and deep learning models predicted relative nutrient-limited yields with low RMSE, and permutation importance identified soil pH as the dominant variable controlling N- and P-limited yields (Ahmed et al., 2024). The same analysis showed that soil N and Zn strongly influenced Zn-limited yield, while spatial trends in K-limited yield emerged along an east-west gradient, revealing distinct fertility constraints for different nutrients. These findings suggest that soil nutrient variables-especially applied N, soil P, Zn, pH and texture-related properties-provide high marginal gains in predictive power and are indispensable components of maize yield models based on soil-climate interactions.

8.2 Influence weight analysis of climate variables

Across diverse modeling frameworks, climate variables frequently emerge as the largest single contributors to interannual maize yield variability. A global meta-analysis using 68 simulation studies for wheat, maize and rice showed that maximum temperature and precipitation significantly affected yield responses, with yields declining by 4.21% per 1 °C increase in maximum temperature but increasing by 0.43% per 1% rise in precipitation (Qin et al., 2023). This quantitative gradient highlights the high negative weight of heat stress and the compensating effect of adequate rainfall in crop-climate response functions. At the global scale, mixed-effects models updating projected yield responses under CMIP6 scenarios indicate that temperature-related stress is a dominant driver of future maize yield losses, with projected global maize declines around 22% by late century under high emissions if adaptation is limited (Li et al., 2025).

Machine-learning-based attribution provides more detailed rankings of individual climate indicators. A hybrid GGCM-Random Forest framework for China’s maize belt found that chilling days, drought indicators and crop pests/diseases were the main factors influencing projected maize yield changes, with relative importance quantified via RF partial-dependence analysis (Li et al., 2023). In a separate process-based and ML study on wheat under future climate scenarios, precipitation explained most yield variability in mid-century high-emission conditions, whereas maximum temperature became the dominant limiting factor under later, more strongly warmed scenarios (El-Mahroug et al., 2025). For site-specific maize prediction with spatio-temporal XGBoost models, precipitation during the juvenile growth phase (May) was identified as the single most important factor