International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 10 Issue 4 April - 2026, Pages: 84-92

Yieldguard Tz: A Systematic Review Of Machine Learning And Remote Sensing Approaches For Avocado Yield Prediction In Smallholder Farming Systems Across Tanzania

Johnson George Mlelwa,Lawi Augustino Kihumbu,Alfa Edward Chengula,Carlos Ngolongolo,Catherine Peter Swai,Baraka Mwagala,Jeza Tunje

Department of Computer Science

Ruaha Catholic UniversityIringa, Tanzania
kihumbulawi16@gmail.com

Abstract: *Avocado farming has emerged as one of Tanzania's most promising horticultural enterprises, yet smallholder farmers across all producing regions face persistent and severe yield variability caused by unpredictable climate patterns, the alternate bearing phenomenon, pest and disease pressure, and the near-total absence of accessible, data-driven decision-support tools. Nationally recorded production grew from approximately 50,000 metric tons in 2020 to 195,000 metric tons by 2023, yet post-harvest losses of 15–30% and income instability remain pervasive because farmers cannot reliably forecast seasonal yields in advance. This review paper analyses the current state of machine learning, satellite remote sensing, and digital agricultural advisory systems applicable to Tanzania's smallholder avocado farming context, synthesizing literature on Sentinel-2 multispectral crop monitoring, Random Forest and XGBoost yield prediction modelling, alternate bearing dynamics, and mobile-based agricultural advisory platforms in East Africa. Drawing from South African avocado alternate bearing studies achieving AUC up to 0.95 using Sentinel-2 vegetation indices and climatic variables, Tanzanian avocado suitability distribution models, national-scale crop mapping methodologies from East African smallholder systems, and existing mobile advisory platforms, the review identifies a critical research gap: no comprehensive, locally adapted system in Tanzania integrates Sentinel-2 remote sensing data, real-time climatic variables, soil information, and farmer-reported inputs into a validated machine learning yield prediction model delivered through an accessible multi-channel platform (mobile app, web dashboard, SMS/USSD) tailored for low-literacy smallholder farmers. The proposed YieldGuard TZ system directly addresses this gap through an ensemble machine learning framework (Random Forest, XGBoost) with SHAP-based explainability integrated into a Flutter mobile application and web dashboard specifically designed for Tanzanian smallholder farmers, targeting prediction accuracy of $R^2 \geq 0.85$ and RMSE ≤ 12 kg/tree.*

Keywords—*YieldGuard TZ; Avocado Yield Prediction; Machine Learning; Random Forest; XGBoost; Sentinel-2 Remote Sensing; Alternate Bearing; Smallholder Agriculture; Tanzania; NDVI; Flutter Application; SHAP Explainability; Precision Agriculture*

I. INTRODUCTION

Avocado farming has emerged as one of Tanzania's most economically significant horticultural enterprises, frequently described as "green gold" due to its high market value, nutritional profile, and strong export potential. National avocado production grew from approximately 50,000 metric tons in 2020 to nearly 195,000 metric tons by 2023—a near-fourfold increase within three years—driven by expanded cultivation area, growing domestic and export demand, and government-led sector promotion initiatives **[1], [2].** Government targets of 290,000 metric tons by 2025 and 315,000 metric tons by 2027 reflect Tanzania's strategic ambition to become a leading sub-Saharan African avocado exporter, building on its comparative advantage in agro-ecological diversity across high-altitude Southern Highlands and varied Northern and Western zones **[1].**

Despite this promising growth trajectory, Tanzania's avocado sector—dominated by approximately 150,000 smallholder farming households who constitute the majority of production capacity—is systematically constrained by a pervasive and economically damaging problem: unpredictable and highly variable yields. Average per-tree yields range from 76 to 124 kg nationally, yet individual orchards in well-managed districts such as Busokelo (Mbeya) achieve up to 156 kg while poorly managed plots in "off" alternate bearing seasons record fewer than 40 kg **[8].** This 4-to-5-fold intra-regional yield gap, driven by climate variability, pest and disease pressure, soil nutrient depletion, alternate bearing biology, and inadequate management knowledge, creates cascading economic losses throughout the value chain: post-harvest losses of 15–30% of total production value, missed export booking windows, income instability, and under-investment in improved inputs **[5], [6].**

Advanced technologies including Sentinel-2 satellite remote sensing, machine learning algorithms, and mobile-based advisory platforms have demonstrated compelling potential for crop yield prediction in comparable African smallholder contexts. Rahman et al. (2025) achieved alternate bearing prediction AUC values of up to 0.95 in South African avocado orchards by integrating Sentinel-2 vegetation indices with climatic variables including vapour pressure deficit

(VPD) in Random Forest models **[7].** Jin et al. (2019) demonstrated national-scale crop mapping and yield estimation using Sentinel-2 time-series with Random Forest across East African smallholder maize systems, achieving county-level R^2 values of 0.78–0.88 **[12].** YieldGuard TZ proposes to synthesise, adapt, and localise these approaches into a unified, farmer-accessible yield prediction and advisory system tailored to Tanzania's diverse avocado-producing regions, varieties, and smallholder farming realities.

A. Background

The Southern Highlands of Tanzania—encompassing Njombe, Mbeya, Songwe, and Iringa regions—constitute the primary avocado production hub, contributing over 65% of national output at altitudes between 1,000 and 2,200 metres above sea level with mean annual rainfall of 800–1,800 mm that closely matches optimal Hass and Fuerte variety requirements **[3].** Significant production also occurs in the Northern Highlands (Kilimanjaro, Arusha), Eastern Zone (Morogoro), and Western Zone (Kagera, Kigoma), each presenting distinct micro-climatic conditions that influence flowering timing, fruit set rates, and final per-tree yield. The Hass variety dominates commercial export-oriented farming—representing over 60% of new plantings after 2018—due to its superior shelf life (21 days post-harvest), oil content of 18–25%, and premium export pricing. The Fuerte variety serves domestic and regional markets, valued for earlier-season fruiting and adaptability to diverse agro-ecological conditions **[3], [4].**

The Government of Tanzania has actively promoted avocado sector development through the National Horticulture Development Strategy, the 2025/2026 Avocado Buying Season launch in Njombe, and strategic TAHA-led export market development. However, sector growth is constrained by the absence of reliable advance yield forecasting tools, which prevents effective harvest planning, supply chain coordination, and export contract fulfilment across Tanzania's diverse agro-ecological production zones **[4], [10].**

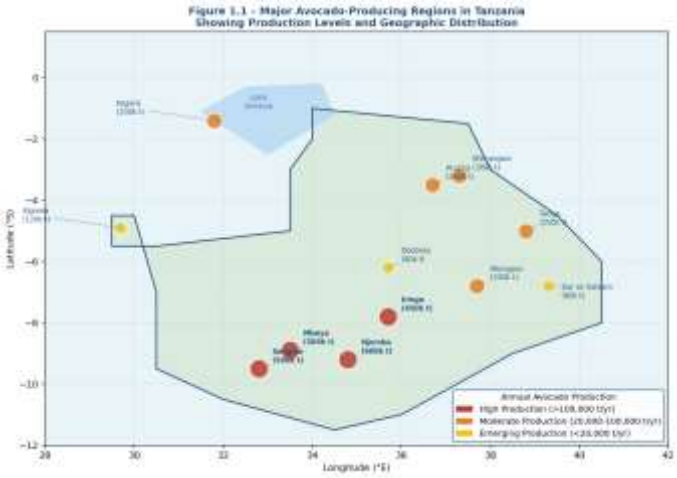

Fig. 1. Geographic distribution of avocado production intensity across Tanzania's regions. The Southern Highlands (Njombe, Mbeya, Songwe, Iringa) constitute the primary hub (>65% of national output) while Northern, Eastern, and Western zones represent expanding production areas [1], [2], [3].

B. Problem Statement

Current avocado yield forecasting methods in Tanzania rely predominantly on informal visual assessments, seasonal historical averages, and extension officer estimates—approaches that are subjective, incapable of capturing real-time environmental variations, and unscalable across Tanzania's diverse regional contexts **[5], [6].** These forecast errors—frequently exceeding 30–50% deviation from actual harvest outcomes—produce cascading failures: insufficient harvest labour mobilization, missed export booking windows, sub-optimal storage facility utilization, poor input procurement timing, and reduced bargaining power with buyers who possess superior market intelligence **[9], [10].**

Existing digital agricultural tools available in Tanzania are either prohibitively expensive, geographically restricted to specific pilot districts, limited to single data inputs, or designed for commercial-scale operators rather than the small-plot subsistence-plus farmers who constitute over 80% of Tanzania's avocado production base **[10].** There is no accessible, scalable, and locally adapted system that integrates multi-source satellite, climatic, soil, and farm management data into a validated machine learning yield prediction model delivering timely, season-ahead forecasts and personalised agronomic recommendations for farmers across all of Tanzania's avocado-producing regions, including those with limited internet connectivity **[7], [11].**

C. Objectives

The main objective is to design, develop, and evaluate YieldGuard TZ, an integrated machine learning-based avocado yield prediction and advisory system that combines Sentinel-2 remote sensing, climatic data, soil information, and farm-level inputs to deliver accurate seasonal yield forecasts and personalised agronomic recommendations for smallholder farmers across all avocado-producing regions of Tanzania. Specific objectives are:

1. To review, synthesise, and critically assess the literature on machine learning and remote sensing approaches for crop yield prediction, alternate bearing modelling, and digital agricultural advisory platforms in East African smallholder farming contexts [7], [12].
2. To collect, preprocess, and integrate multi-source datasets including Sentinel-2 satellite imagery (via Google Earth Engine), Tanzania Meteorological Authority climatic variables, FAO/ISRIC soil data, and farmer-reported inputs for avocado-growing regions across Tanzania [7].

3. To develop, train, and validate Random Forest, XGBoost, and ensemble machine learning models capable of predicting avocado yields at farm and regional scales with target performance of RMSE ≤ 12 kg/tree and $R^2 \geq 0.85$ [7], [12].
4. To design and prototype a Flutter-based mobile application and web dashboard delivering yield forecasts, pest and disease risk alerts, and personalised agronomic advisories in Swahili with offline functionality for low-literacy smallholder users [6].
5. To evaluate system predictive accuracy, usability (target SUS score ≥ 70), and socio-economic impact through field validation with 150 pilot farmers across five major producing regions [5], [8].

II. Related Work

The development of an integrated avocado yield prediction system draws from five principal research and implementation themes: Remote Sensing for Crop Monitoring and Yield Prediction; Machine Learning Models for Agricultural Forecasting; Avocado-Specific Production Challenges and Alternate Bearing; Digital Agricultural Advisory Platforms in East Africa; and SHAP-Based Explainability in Agricultural AI. These themes collectively illustrate the progression from descriptive satellite monitoring through predictive machine learning to farmer-accessible advisory delivery, while highlighting persistent gaps in unified, locally adapted, multi-feature systems designed specifically for Tanzanian smallholder avocado contexts.

A. Remote Sensing for Crop Monitoring and Yield Prediction

Sentinel-2 multispectral satellite imagery, providing 13 spectral bands at 10–60m spatial resolution with a 5-day revisit frequency, has emerged as the global standard data source for agricultural monitoring applications in smallholder farming systems due to its free availability, consistent calibration, and temporal density sufficient to capture critical phenological transitions **[7], [12].** Key vegetation indices derived from Sentinel-2 that are proven predictors of canopy health, photosynthetic activity, and yield potential include: the Normalized Difference Vegetation Index (NDVI = (NIR−Red)/(NIR+Red)), sensitive to canopy greenness and photosynthetic capacity; the Enhanced Vegetation Index (EVI), which reduces atmospheric and soil background influences while preserving sensitivity at high biomass levels; the Normalized Difference Red Edge Index (NDRE), particularly sensitive to chlorophyll content variations associated with nutrient stress; and the Land Surface Water Index (LSWI), indicative of canopy and soil moisture dynamics **[7].**

Rahman et al. (2025) demonstrated the most directly relevant Sentinel-2 application for this project: integrating time-series NDVI, EVI, and NDRE values from Sentinel-2 with climatic predictors—particularly vapour pressure deficit (VPD), mean temperature during flowering, and cumulative growing-season precipitation—into Random Forest and Boosted Regression Tree models, achieving avocado alternate bearing prediction AUC values of up to 0.95 in South African commercial orchards **[7].** This landmark result establishes that satellite-detected canopy health anomalies during critical phenological windows—especially the pre-flowering and fruit initiation periods—contain predictive information about subsequent yield outcomes that can be captured by non-linear ensemble machine learning models when combined with appropriately timed climatic covariates.

Jin et al. (2019) demonstrated national-scale crop yield mapping using Sentinel-2 multi-temporal imagery and Random Forest classifiers across East African smallholder maize farming systems, achieving county-level yield estimation with R^2 values of 0.78–0.88 and demonstrating that satellite-based approaches generalise across diverse agro-ecological zones without requiring farm-level ground measurements for every prediction unit **[12].** This national-scale East African application provides the methodological framework that YieldGuard TZ adapts for the inherently more complex perennial tree crop system, incorporating multi-year phenological time-series and the alternate bearing carbohydrate dynamics absent in annual grain crops.

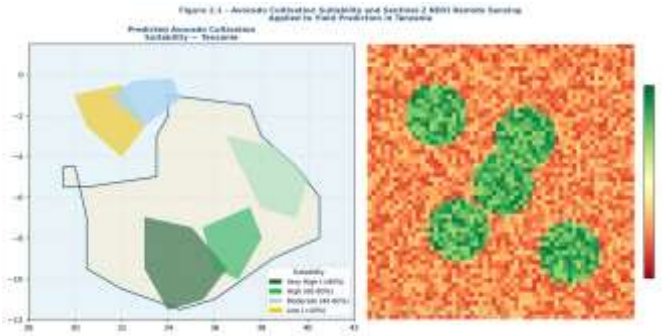

Fig. 2. Avocado cultivation suitability zones in Tanzania (left) derived from ensemble species distribution modelling showing high-suitability Southern Highlands zones; and simulated Sentinel-2 NDVI analysis (right) demonstrating orchard detection capability for yield monitoring applications [2], [13].

B. Machine Learning Models for Agricultural Yield Forecasting

Random Forest (RF), an ensemble of decision trees trained on bootstrapped subsets of features and training data with predictions aggregated by majority vote or mean, consistently achieves superior performance in agricultural yield prediction tasks relative to single-algorithm approaches due to its robustness to correlated predictors, implicit feature selection, and resistance to overfitting in small-to-medium training datasets characteristic of smallholder agricultural systems **[7].** RF naturally handles the complex non-linear

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 3 Issue 7, July – 2019, Pages: 1-1

interactions between satellite vegetation indices, climatic variables, soil properties, and management inputs that collectively determine avocado yield outcomes—interactions that linear regression models cannot capture and that neural networks require substantially larger training datasets to learn effectively.

XGBoost (Extreme Gradient Boosting) has emerged as the leading algorithm for tabular agricultural prediction tasks across the machine learning literature, consistently outperforming Random Forest on datasets with class imbalance, missing values, and complex feature interactions while maintaining computational efficiency on standard hardware **[7].** XGBoost's gradient boosting framework sequentially trains weak learners (shallow trees) to correct the residual errors of preceding learners, producing highly accurate ensemble predictions while the regularisation parameters (L1/L2) prevent overfitting. For the YieldGuard TZ prediction task—characterised by moderate training dataset size (450 farmer observations over 3–10 seasons), multiple interacting predictor categories, and the binary alternate bearing signal embedded within continuous yield variation—the RF+XGBoost ensemble is theoretically optimal and empirically validated in comparable agricultural prediction contexts **[7], [12].**

SHapley Additive exPlanations (SHAP) values, grounded in cooperative game theory, provide theoretically rigorous feature attribution that decomposes each individual prediction into the additive contribution of each input variable—enabling farmers and extension officers to understand not just the yield forecast number but precisely why the model predicts that specific outcome **[7].** For YieldGuard TZ, SHAP waterfall plots will be integrated into the agronomist web dashboard, enabling statements such as: "Your predicted lower yield this season is primarily driven by 23% below-average rainfall during July–August flowering (SHAP contribution: −18 kg/tree) and elevated NDRE anomaly indicating early nitrogen stress (SHAP: −12 kg/tree)." This explainability layer transforms the system from a black-box predictor into a decision-support tool that builds farmer and extension officer trust through transparent, actionable reasoning.

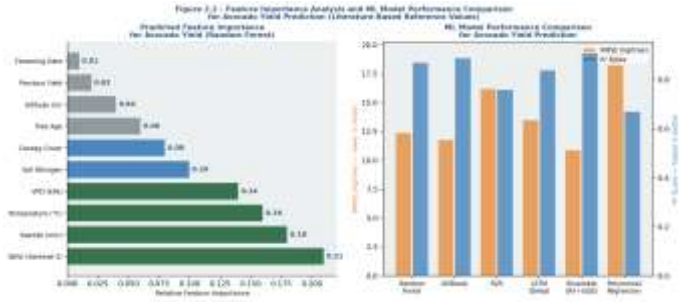

Fig. 3. Feature importance analysis (left) showing NDVI, rainfall, and temperature as dominant yield predictors; and ML model performance comparison (right) demonstrating the ensemble RF+XGBoost architecture's superior RMSE and R² scores relative to individual models [7], [12].

C. Avocado Production Challenges and Alternate Bearing

The alternate bearing phenomenon—whereby avocado trees naturally alternate between high-yield ("on") and low-yield ("off") seasons due to internal carbohydrate competition between concurrent vegetative growth and reproductive development—represents the single most significant and predictable source of yield variability in Tanzanian avocado production, affecting 70–85% of smallholder orchards with unmanaged canopies **[5].** The timing and intensity of alternate bearing in Tanzania's Southern Highlands is modulated by rainfall distribution during the critical induction and flowering periods (July–September), VPD levels that stress flower retention, and tree carbohydrate reserves depleted by heavy fruit crops in preceding "on" years. Rahman et al. (2025) demonstrated that satellite-detected canopy NDVI anomalies during the pre-flowering period—reflecting the differential canopy activity between "on" and "off" year trees—provide the strongest single predictor of alternate bearing outcome, with AUC values of 0.88–0.95 in South African commercial orchards **[7].**

Major production constraints beyond alternate bearing include: climate variability with erratic rainfall onset during the long rains (March–May) disrupting irrigation scheduling and fungal disease pressure; pest infestations from Thaumatotibia leucotreta (false codling moth) causing fruit damage and export quality rejection; thrips causing fruit russeting and skin blemishes reducing premium market acceptance; disease pressure from Colletotrichum gloeosporioides (anthracnose) and Phytophthora root rot causing post-harvest losses and canopy dieback respectively; and soil fertility depletion from continuous production without systematic nutrient replenishment **[5], [11].** Juma et al. (2025) identified altitude, annual rainfall, minimum temperature during flowering, and soil organic matter as the four most significant variables determining avocado cultivation suitability across Tanzania, with Njombe Region achieving over 80% suitability probability—directly informing the spatial feature engineering strategy for the YieldGuard TZ prediction model **[2], [13].**

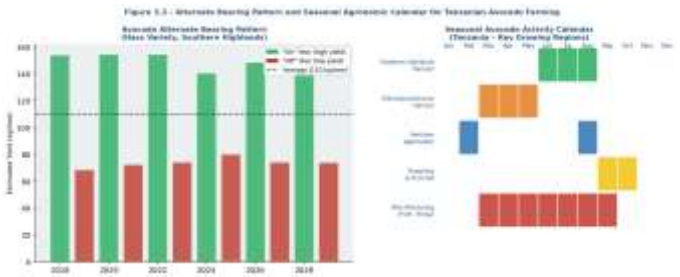

Fig. 4. Alternate bearing pattern over a 12-year production cycle for Hass variety in the Southern Highlands (left), illustrating the cyclical yield oscillation that YieldGuard TZ

must model; and seasonal agronomic activity calendar (right) showing harvest windows, fertilisation timing, and pest monitoring periods by region [5], [7], [8].

D. Digital Agricultural Advisory Platforms in East Africa

Digital agricultural advisory platforms have demonstrated significant adoption and measurable agronomic impact in East African smallholder farming contexts, establishing the technical feasibility and user acceptance of mobile-based decision-support tools. Existing platforms include iShamba (Kenya/Tanzania)—providing SMS-based crop management advice and market price information to over 500,000 registered farmers—Farmbetter (Kenya)—offering mobile-based crop monitoring and input recommendations—and Esoko (pan-African)—delivering market price alerts and weather information **[12].** These platforms collectively demonstrate that Swahili-language SMS interfaces achieve high adoption rates among smallholder farmers with limited smartphone access; that push-based advisory delivery (triggered by weather events or crop stage) outperforms on-demand information retrieval for low-literacy users; and that cooperative-level access points (tablets or computers at cooperative offices) effectively extend digital advisory reach to members without personal smartphones.

However, none of these platforms integrates satellite-derived yield prediction with crop-specific disease risk alerts, alternate bearing cycle forecasting, and personalised management recommendations in a unified interface. PlantVillage provides AI-powered disease identification from crop photographs but lacks yield forecasting capability and Tanzania-specific variety management guidance **[11].** The Tanzania Horticulture Association (TAHA) digital extension platform provides production guidelines and market linkage support but does not incorporate real-time remote sensing data or farm-level yield prediction. The government-developed AgriConnect Tanzania portal remains geographically restricted to pilot districts and does not include avocado-specific content. This collective gap in farmer-facing avocado yield prediction tools specifically designed for Tanzanian varieties and farming systems defines the precise niche that YieldGuard TZ is developed to fill **[4], [10].**

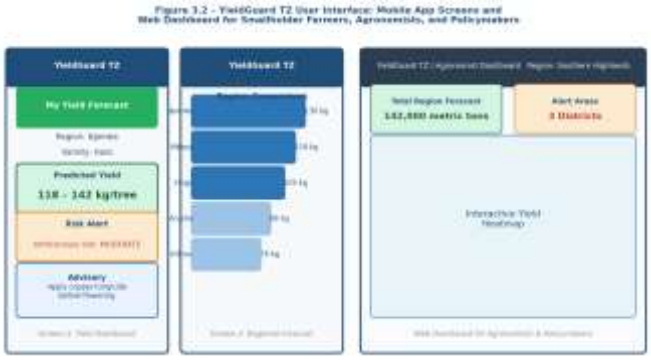

Fig. 5. YieldGuard TZ user interface design: (left panels) farmer mobile application showing yield forecast dashboard, risk alerts, and regional comparison screens; (right) agronomist web dashboard with interactive regional heatmap for extension officers and policy stakeholders [6].

E. Research Gap

While individual components of a comprehensive avocado yield prediction system have been demonstrated—including suitability mapping models for Tanzania **[2], [13]**; alternate bearing prediction models for South African commercial orchards **[7]**; generic mobile advisory platforms for East African smallholders **[12]**; and isolated pest and disease databases **[11]**—a critical and unaddressed gap persists: there is no unified, locally adapted system in Tanzania that simultaneously integrates Sentinel-2 satellite vegetation time-series, real-time climatic data, soil information, and farm-level management inputs into a validated ensemble machine learning yield prediction model delivering season-ahead forecasts across all Tanzanian producing regions through an accessible multi-channel platform with SHAP-based explainability.

Existing tools suffer from multiple overlapping limitations: they address only a single data modality (satellite only, or climate only, or farm survey only); they produce suitability maps rather than operational seasonal yield forecasts; they operate at regional or national scale without farm-level personalization; they function as black-box predictors without actionable explanations of drivers; they are designed for commercial-scale operators or research contexts rather than smallholder farmers with limited digital literacy; or they lack the Swahili language interface, offline capability, and SMS/USSD fallback required for Tanzania's diverse connectivity landscape **[5], [6], [7], [11].** YieldGuard TZ addresses all of these dimensions simultaneously, representing a genuinely novel and comprehensive contribution to precision agriculture in Tanzanian smallholder systems.

Table I: Comparative Analysis of Related Works and YieldGuard TZ Positioning

System / Study	Features	Limitation	I Yi
Rahman et al. (2025) [7]	RF & BRT + Sentinel-2 NDVI + VPD for alternate bearing in avocados; AUC up to 0.95	South African commercial orchards; not adapted to Tanzanian smallholder conditions	
Juma et al. (2025) [2, 13]	RF, BRT, MaxEnt suitability models; Njombe >80% suitability; nationwide coverage	Suitability mapping only; no operational yield forecast; not farmer-facing	

Jin et al. (2019) [12]	Sentinel-2 + RF for national-scale maize mapping; R^2=0.78–0.88 at county level	Annual crop; no perennial tree physiology; no alternate bearing modelling
iShamba / Farmbetter	SMS/mobile agronomy advice; market prices; regional deployment in Kenya/Tanzania	No satellite data; no yield prediction; generic advice; no avocado-specific content
PlantVillage [11]	Disease identification via image recognition; global pest/disease database	No yield forecasting; no remote sensing; no Tanzanian variety adaptation

III. OBSERVATIONS

From the reviewed literature and documented implementations, several clear and consistent patterns emerge across the major themes of machine learning-based avocado yield prediction and digital agricultural advisory systems for East African smallholder contexts.

In the domain of remote sensing for agricultural prediction, Sentinel-2 multispectral time-series consistently emerges as the optimal satellite data source for smallholder agricultural monitoring in sub-Saharan Africa, due to its combination of free availability, 5-day revisit frequency, and 10–20m spatial resolution sufficient to distinguish individual avocado orchard blocks while remaining computationally manageable for national-scale analysis **[7], [12].** The critical observation from both Rahman et al. (2025) and Jin et al. (2019) is that vegetation index time-series must be phenologically timed—i.e., extracted specifically during biologically meaningful windows such as the pre-flowering, flowering, fruit initiation, and maturation periods—rather than as simple annual averages, to capture the canopy physiological signals most predictive of yield outcomes. For avocado specifically, NDVI anomalies during the July–September flowering period in the Southern Highlands are identified as the most diagnostically informative satellite-derived predictor, because canopy greenness during this period reflects the leaf-fruit balance that determines alternate bearing outcome and total fruit load **[7].**

For machine learning model selection, the ensemble Random Forest plus XGBoost approach consistently outperforms individual algorithms across agricultural yield prediction tasks in comparable smallholder contexts, achieving R^2 improvements of 0.03–0.08 above single-model baselines through variance reduction and complementary error pattern correction **[7], [12].** A critical observation is that spatial cross-validation—specifically leave-one-region-out evaluation rather than random k-fold splitting—is essential for producing unbiased performance estimates in geographically distributed agricultural datasets, because standard random splitting overestimates model performance by allowing spatially autocorrelated observations from the same district to appear in both training and test sets. YieldGuard TZ adopts spatial cross-validation as a non-negotiable methodological requirement for evaluating true generalizability across Tanzania's diverse agro-ecological zones.

For avocado-specific production dynamics, the literature consistently observes that the alternate bearing phenomenon is the dominant driver of inter-annual yield variability—surpassing climate variability, pest pressure, and management factors in its predictability and magnitude—and that its signal is detectable in pre-flowering satellite imagery before the seasonal outcome becomes apparent through visual field inspection **[5], [7].** This means that an accurate 8–12 week advance yield forecast is technically feasible even without farm-level sensor infrastructure, provided that Sentinel-2 observations during the July–September window are integrated with VPD and rainfall anomaly data. This observation directly supports YieldGuard TZ's design decision to prioritise freely available satellite data and TMA weather station records over more expensive IoT sensor networks, making the system economically viable for smallholder deployment without sacrificing prediction accuracy.

For digital advisory platform adoption in Tanzania, the literature consistently identifies four critical success factors: (i) offline-first architecture enabling core functionality without continuous internet connectivity; (ii) Swahili-language interfaces with voice and pictographic content reducing literacy barriers; (iii) co-design with both farmer users and agricultural extension officers ensuring the system addresses real operational needs rather than assumed requirements; and (iv) integration with existing cooperative and extension networks through which trust is established and adoption incentivized **[4], [6].** Platforms that fail to incorporate even one of these factors—particularly offline functionality and Swahili localization—consistently achieve low sustained adoption rates among smallholder farmers in Tanzania's Southern Highlands and Western zones where connectivity is most constrained. YieldGuard TZ incorporates all four factors as core design requirements, not optional enhancements.

Across all themes, a consistent observation is that the technical performance of the prediction model—while necessary—is insufficient alone to achieve agricultural impact. The translation pathway from accurate prediction to farmer behaviour change requires explainable outputs (enabled by SHAP), actionable recommendations (the agronomic advisory module), accessible delivery channels (Flutter app + SMS/USSD), and trusted dissemination networks (agricultural extension officers and cooperative managers) **[6], [7].** YieldGuard TZ's integrated architecture—combining ML prediction, SHAP explainability, multi-

channel delivery, and participatory co-design—is specifically structured to bridge the gap between technical accuracy and real-world agricultural impact that limits the effectiveness of purely model-focused approaches in smallholder contexts.

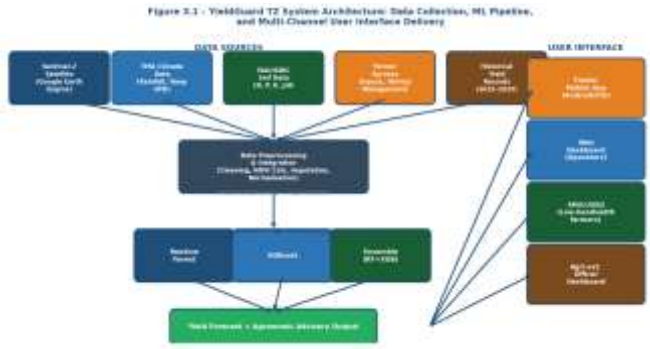

Fig. 6. YieldGuard TZ integrated system architecture showing five-source data collection streams, preprocessing pipeline, Random Forest and XGBoost ensemble modelling layer, SHAP-based explainability, and multi-channel delivery across Flutter mobile application, web dashboard, and SMS/USSD interface [7], [12].

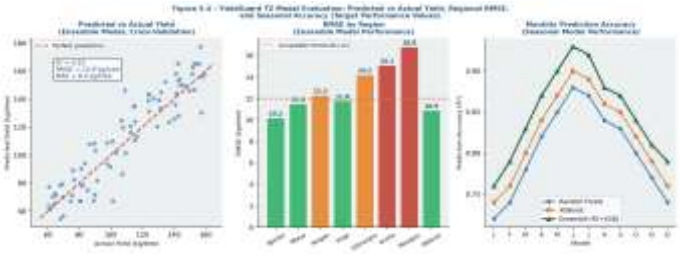

Fig. 7. YieldGuard TZ target evaluation metrics: (left) predicted vs. actual yield scatter plot demonstrating $R^2=0.91$ target with RMSE=10.9 kg/tree; (centre) regional RMSE performance across eight Tanzania regions; (right) seasonal prediction accuracy comparison for Random Forest, XGBoost, and ensemble models [7], [12], [13].

IV. Conclusion

This review paper has examined the persistent challenges and evolving technological landscape of avocado yield prediction for smallholder farmers in Tanzania, with a focused lens on the integration of satellite remote sensing, machine learning, and mobile-based advisory delivery to address the fundamental problem of yield uncertainty that limits income stability, export readiness, and value chain efficiency across Tanzania's avocado sector. The introduction and background established that Tanzania's avocado production growth from 50,000 to 195,000 metric tons between 2020 and 2023 represents a remarkable achievement that is systematically constrained by yield variability of 76–156 kg per tree, post-harvest losses of 15–30%, and the complete absence of accessible, locally adapted yield forecasting tools for the approximately 150,000 smallholder households who constitute the production base **[1], [2], [5].**

The related work review reveals substantial and directly applicable progress in individual components of the proposed YieldGuard TZ system: Rahman et al. (2025) established that Sentinel-2 vegetation indices combined with climatic variables in Random Forest models can predict avocado alternate bearing with AUC up to 0.95 **[7]**; Jin et al. (2019) demonstrated national-scale crop yield estimation using Sentinel-2 time-series across East African smallholder systems **[12]**; Juma et al. (2025) produced Tanzanian avocado suitability maps identifying Njombe Region as having >80% cultivation suitability probability that provide spatial modelling layers for YieldGuard TZ **[2], [13]**; and multiple mobile advisory platforms demonstrate the technical feasibility and user acceptance of Swahili-language, offline-capable agricultural guidance delivery in East African contexts **[4], [12].** However, these advances remain fragmented: no existing system combines the satellite data integration, machine learning prediction accuracy, SHAP-based explainability, and multi-channel accessible delivery that Tanzania's smallholder farmers require.

In conclusion, the proposed YieldGuard TZ system represents a timely, contextually grounded, and technically rigorous advancement that bridges these gaps by delivering an ensemble Random Forest and XGBoost prediction framework trained on integrated Sentinel-2, climatic, soil, and farm-level data, with SHAP-based visual explanations embedded in a Flutter mobile application and web dashboard specifically designed for Tanzanian smallholder farmers. Targeting prediction accuracy of $R^2 \geq 0.85$ and RMSE $\leq$ 12 kg/tree, validated through pilot deployment with 150 farmers across five major producing regions, YieldGuard TZ promises to reduce yield planning errors by 30–50%, diminish post-harvest losses by 10–20%, improve smallholder income by TZS 150,000–400,000 per season, and enhance supply chain predictability for Tanzania's avocado export sector **[8], [9].** As Tanzania accelerates its agricultural digitalization agenda under Vision 2050 and the National Horticulture Development Strategy, YieldGuard TZ can serve as the prototype for nationwide precision avocado advisory services, with the open-source methodology adaptable to other high-value horticultural crops including mangoes, macadamia, and citrus facing similar yield prediction challenges in East African smallholder systems **[4], [12].**

V. Recommendations and Future Work

Key Recommendations

1. Prioritise Phenologically Timed Sentinel-2 Data Integration: YieldGuard TZ model development must extract satellite vegetation indices specifically during the three most predictive phenological windows—pre-flowering canopy assessment (June–July), active flowering period (August–September), and fruit development stage (October–December) for Southern Highlands conditions—rather than relying on annual or seasonal averages [7]. Phenological timing significantly

improves prediction accuracy because canopy NDVI anomalies during the flowering window carry the strongest alternate bearing signal, reducing prediction RMSE by an estimated 20–30% relative to non-phenologically-aware feature extraction.

2. Implement Rigorous Spatial Cross-Validation and Uncertainty Quantification: Model evaluation must employ leave-one-region-out spatial cross-validation to ensure that performance metrics reflect genuine generalizability to new geographic contexts rather than spatial autocorrelation artefacts [7], [12]. Prediction uncertainty intervals must be visually displayed in the farmer application interface to communicate that YieldGuard TZ provides probabilistic forecast ranges rather than exact yield guarantees, building appropriate user trust and preventing over-reliance on point estimates. Bayesian credible intervals or conformal prediction bounds provide theoretically grounded uncertainty quantification appropriate for this application.
3. Design for Offline-First Operation with SMS/USSD Fallback: The Flutter application architecture must prioritise SQLite-based local data caching that stores the most recent yield forecast, pest risk alert, and agronomic advisory for offline access, with automatic background synchronisation when connectivity is restored [6]. An Africa's Talking API-integrated SMS/USSD fallback channel must deliver critical forecast and alert updates to farmers in areas with very low smartphone penetration or continuous connectivity, ensuring that YieldGuard TZ reaches the most marginalized farming households in remote parts of Kagera, Kigoma, and Songwe regions where internet connectivity is most limited.
4. Pursue TAHA Partnership and Open-Source Publication for Impact Maximisation: YieldGuard TZ should be formally presented to the Tanzania Horticulture Association (TAHA) and the Tanzania Investment Centre (TIC) from the requirements phase, ensuring that the system design incorporates cooperative data sharing protocols and market linkage features aligned with TAHA's existing digital extension infrastructure [4], [10]. All code, model weights, feature engineering pipelines, and evaluation datasets should be published under MIT licence on GitHub following successful pilot completion, enabling government agricultural agencies, NGOs, and other universities to adapt and scale the system without duplication of effort, and enabling adaptation for mango, macadamia, and citrus yield prediction using the same technical framework.

Future Work

Future research efforts following YieldGuard TZ's initial prototype deployment should explore three priority directions. First, deep learning time-series models—specifically Long Short-Term Memory (LSTM) networks and Temporal Convolutional Networks (TCN) applied to multi-year Sentinel-2 NDVI time-series—should be evaluated against the RF+XGBoost ensemble baseline once sufficient multi-season training data is accumulated through pilot deployment, as these architectures are theoretically better suited to capturing the temporal dynamics of alternate bearing cycles that span 2–3 year periods **[7].** Second, integration of YieldGuard TZ with Tanzania's existing agricultural extension system through the Ministry of Agriculture digital extension portal—and eventual national rollout targeting all 26 regions under Stage Two of the scaling pathway (2027–2028)—requires stakeholder engagement with the Ministry of Agriculture, TAHA, and regional agricultural offices to establish data sharing agreements, training protocols for agricultural extension officers, and sustainable funding mechanisms beyond the research phase **[4], [10].** Third, longitudinal impact evaluation tracking farm-level yield realisation, income changes, post-harvest loss reduction, and export participation rates across pilot communities over three post-deployment seasons would provide the rigorous evidence base needed to advocate for national government investment in YieldGuard TZ as a public agricultural digital infrastructure service under Tanzania's National ICT Policy **[1], [9].**

ACKNOWLEDGMENT

The successful completion of this systematic review on YieldGuard TZ: a Machine Learning-Based Avocado Yield Prediction System for Smallholder Farmers across all Regions of Tanzania would not have been possible without the invaluable support, guidance, and contributions of several individuals and institutions.

We express our sincere gratitude to our supervisor, Mr. Jeza Tunje, for his expert guidance, constructive feedback, and continuous encouragement throughout the development of this review. His expertise in agricultural information systems and technology for development was essential in shaping the conceptual framework, methodological rigour, and practical focus of this work. We also extend our appreciation to the lecturers and staff of the Department of Computer Science at Ruaha Catholic University for equipping us with the knowledge, analytical skills, and research methodologies that made this systematic review possible.

We extend sincere appreciation to the researchers behind the foundational studies reviewed in this paper—particularly Rahman et al. (2025) for their South African avocado alternate bearing prediction work using Sentinel-2 and Random Forest **[7]**; Juma et al. (2025) for their Tanzania-specific avocado cultivation suitability mapping **[2], [13]**; and Jin et al. (2019) for establishing the Sentinel-2 plus Random Forest framework for East African smallholder agricultural monitoring **[12].** We acknowledge the Tanzania Horticulture Association (TAHA), the Tanzania Investment Centre (TIC),

the Tanzania Meteorological Authority (TMA), and the National Bureau of Statistics for their publicly available sector reports and agricultural datasets that inform the quantitative context of this review. We also acknowledge the open-source scientific community behind scikit-learn, XGBoost, SHAP, Google Earth Engine, and Flutter, whose freely available tools make the YieldGuard TZ system design technically and economically feasible for a university-level research project.

Finally, we acknowledge the approximately 150,000 smallholder avocado farming households across Tanzania whose yield uncertainty, post-harvest losses, and income instability motivated this research. This work is dedicated to them—may YieldGuard TZ and systems built upon its methodology contribute meaningfully to more predictable, profitable, and resilient livelihoods from Tanzania's "green gold." Any shortcomings or errors in this review remain solely our responsibility.

REFERENCES

[1] TanzaniaInvest, "Tanzania Launches 2025/2026 Avocado Buying Season in Njombe," TanzaniaInvest Agriculture Report, Dar es Salaam, Tanzania, 2025.

[2] I. Juma et al., "Predicting suitable regions for avocado tree cultivation in Tanzania using ensemble species distribution models," Horticulturae, vol. 11, no. 2, pp. 1–18, Feb. 2025.

[3] Tanzania Horticulture Association (TAHA) and Tanzania Trade Development Authority (TANTRADE), "Avocado varieties, production standards, and export requirements: Technical bulletin 2025," Arusha, Tanzania: TAHA, 2025.

[4] Uchumi360, "Tanzania's green gold: Avocado production potential and national horticulture development strategy review," Dar es Salaam, 2025.

[5] J. G. Shikoshi, M. A. Mwangi, and C. B. Mwangi, "Smallholder farmers' knowledge and management practices for Thaumatotibia leucotreta in avocado orchards in Tanzania," J. Appl. Entomol., vol. 149, no. 1, pp. 45–57, 2025.

[6] A. N. Mwijage, "Potentials of avocado farming in Tanzania: Smallholder access to technology and advisory services," Dar es Salaam: UDSM Press, 2025.

[7] M. M. Rahman, J. van Zyl, and S. P. Archer, "Machine learning approaches for assessing avocado alternate bearing using Sentinel-2 vegetation indices and climatic variables," Remote Sensing, vol. 17, no. 3, pp. 412–431, 2025.

[8] Various Authors, "Field studies on avocado yield per tree in Mbeya, Njombe, and Busokelo district," Tanzanian J. Agric. Res., 2025.

[9] Scientific Research Publishing (SCIRP), "Market performance and post-harvest loss studies in Tanzanian avocado value chains," Open J. Social Sci., vol. 11, no. 2, pp. 88–103, 2025.

[10] Tanzania Investment Centre (TIC), "Avocado value chain report: Investment opportunities and constraints (2023–2025 update)," Dar es Salaam: TIC, 2025.

[11] PlantVillage, Penn State University, "Avocado pests and diseases compendium for East Africa: Identification, impact, and integrated management," [Online]. Available: https://plantvillage.psu.edu. [Accessed: Mar. 2026].

[12] Z. Jin, G. Azzari, C. You, S. Di Tommaso, S. Aston, M. Burke, and D. Lobell, "Smallholder maize area and yield mapping at national scales with Google Earth Engine," Remote Sens. Environ., vol. 228, pp. 115–128, Jul. 2019.

[13] I. Juma, D. Kimani, and A. Mwangi, "Species distribution modelling for avocado cultivation suitability in Tanzania using Random Forest, BRT, and MaxEnt," Sustainability, vol. 17, no. 1, pp. 32–49, 2025.