EEET ECOLOGICAL ENGINEERING & ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology, 2025, 26(7), 113–125
https://doi.org/10.12912/27197050/205069
ISSN 2719–7050, License CC-BY 4.0

Received: 2025.04.25
Accepted: 2025.05.30
Published: 2025.06.15

Integrating random forest and irrigation management in geographic information systems-based land suitability and rice productivity modeling in tropical landscapes

Samsu Arif[1,2*], Sakka[1], Nurmiaty[3], Rahmad[3], Erfan Syamsuddin[1], Muh. Farid Wajedy[1], Andri Moh. Wahyu Laode[1], Syamsuddin[1]

[1] Department of Geophysics, Faculty of Mathematics and Natural Science, Hasanuddin University, Makassar, 90245, Indonesia
[2] Research and Development Center for Regional Development and Spatial Information, Hasanuddin University, Makassar, 90245, Indonesia
[3] Pangkep State Polytechnic of Agriculture, Pangkajene Kepulauan Regency, Indonesia
* Corresponding author e-mail: samsu_arif@unhas.ac.id

ABSTRACT

This study was conducted in Barru Regency, Indonesia, a region characterized by diverse topography, high agricultural potential, and environmental constraints. It evaluated the predictive performance of the random forest (RF) machine learning algorithm in FAO-based land suitability classification and rice productivity estimation, integrating geographic information systems (GIS) and technical irrigation as a novel managerial variable. Using 12 GIS-derived soil parameters, the RF model achieved high accuracy (0.95 ± 0.01) for land suitability classification via cross-validation. However, its productivity prediction yielded a low R^2 (0.32 ± 0.04) with 530 filtered samples, likely due to data complexity. The low R^2 value indicates that unmodeled factors, such as agricultural management practices significantly impact productivity beyond the environmental and irrigation variables captured in the model. Including irrigation data improved R^2 by 6%, raising it to 0.38 ± 0.06, highlighting the importance of managerial factors in tropical regions with limited infrastructure. Sensitivity analysis identified slope, cation exchange capacity (CEC), and soil depth as key for land suitability, while slope, potassium, and CEC influenced productivity prediction. The land suitability map showed that class S3 (marginally suitable) dominates (34,836.62 ha), followed by not suitable (N: 8,303.68 ha) and moderately suitable (S2: 2,722.62 ha), with Barru Sub-district having the largest S2 area (1,181.16 ha). These findings suggest that enhancing irrigation infrastructure and improving soil conditions can support better land management strategies. Despite limitations in productivity prediction, the integration of managerial variables represents an innovative approach, improving model performance and practical applications.

Keywords: GIS, random forest, land suitability, rice productivity, precision agriculture, tropical constrained land, machine learning.

INTRODUCTION

Rice production in tropical regions like South Sulawesi, Indonesia, is critical for global food security but is hindered by poor drainage, low soil fertility, and complex topography (Shin et al., 2020; Limpo et al., 2022). Integrating technical irrigation as a novel managerial variable can address these constraints, enhancing land suitability and productivity (Khan et al., 2022). Machine learning and data-driven approaches further improve resource management, enabling sustainable yield increases in these challenging environments to support local and global food needs (Shoaib et al., 2023). Recent advancements in GIS and ML have significantly improved predictive modeling, offering critical data-driven insights for sustainable agricultural planning. Among ML techniques, RF stands out for its ability to model complex, nonlinear relationships across diverse environmental variables, making it especially suitable for the heterogeneous conditions of tropical agriculture (Gbode et al., 2025; Rashid et al., 2021). Its strength lies in handling high-dimensional data and assessing feature importance through permutation importance, enhancing its effectiveness in land suitability assessments and crop yield predictions (Shahhosseini et al., 2021). Empirical studies have shown RF's success in integrating key land characteristics, such as slope, soil depth, and nutrient levels (e.g., nitrogen and phosphorus), to forecast agricultural outcomes and inform tailored land management strategies (Tziolas et al., 2021).

However, the application of RF in tropical regions remains limited. The unique environmental variability and infrastructural challenges in these regions require more than just biophysical parameters; managerial factors such as irrigation and fertilization are often overlooked despite their importance. (Adugna et al., 2022; David et al., 2022; Fei et al., 2023; Taghizadeh-Mehrjardi, Schmidt, et al., 2020). In areas like Barru Regency, South Sulawesi, where seasonal flooding, inadequate drainage, and complex terrain prevail, ignoring such inputs may reduce ecological relevance and predictive accuracy. Moreover, conventional model validation approaches often overlook spatial autocorrelation, leading to overly optimistic performance estimates and reducing generalizability to real-world applications (Wadoux et al., 2020), his study mitigated spatial autocorrelation using 5-fold cross-validation and Out-of-Bag (OOB) scoring to ensure more reliable and generalizable model performance.

To address these gaps, this study investigates the potential of integrating environmental and managerial variables into Random Forest models to improve predictions of land suitability and rice productivity in tropical constrained landscapes. This study aims to:

- evaluate the predictive accuracy of Random Forest for FAO-based land suitability classification and rice yield estimation in Barru Regency using 12 GIS-derived soil parameters and technical irrigation as a novel managerial input.
- assess the relative importance of these variables through sensitivity analysis to identify key drivers of land suitability and productivity.
- develop integrated land suitability mapping to inform targeted agricultural interventions for enhanced productivity in infrastructure-limited tropical regions. These objectives are expected to provide novel insights into the role of managerial factors in enhancing predictive performance and offer a scalable approach for precision agriculture in tropical regions.

METHODS AND MATERIALS

Study area

Barru Regency, located in South Sulawesi, Indonesia (4°08'–4°31' S, 119°36'–119°48' E), spans 1,174.72 km² across seven sub-districts and features diverse topography ranging from coastal zones to elevations of 1,400 m (Figure 1). With 62% of the land used for agriculture, rice is the dominant crop, followed by maize and soybean. The region experiences a tropical climate, with annual rainfall between 2.000 and 3.000 mm during the wet season (November-April) and temperatures ranging from 25 °C to 32 °C. Soils include lowland inceptisols and hilly ultisols, both of which are limited by low organic matter content and high leaching potential.

Field data collection

Field data collection was conducted from November 2023 to April 2024, involving agricultural extension officers from the Barru Regency Agriculture and Food Security Office to measure rice productivity and determine the status of irrigated (technical and non-technical) rice fields across Barru Regency. A total of 530 sampling points across rice field locations were selected using the stratified random sampling method to ensure representation across seven sub-districts, considering farmer group distribution, topographic variations obtained from DEMNAS data, and irrigation coverage. These points were specifically used to collect data on rice productivity (tons/ha) through farmer interviews and direct yield measurements, as well as to record the irrigation status of rice fields. The irrigation status was categorized as technical (managed by engineered systems such as canals or sprinklers, coded as 1) or non-technical (rainfed or traditional systems, coded as 0). Irrigation status was verified using the 2024 irrigation coverage map from the Barru Regency Agriculture Office and field observations at each sampling point. Data on land characteristics (Figure 2: (a) surface rock, (b) slope gradient, (c) soil depth, (d) texture, and (e) irrigated rice fields) and soil fertility parameters (Table 2 and Figure 3: CEC, organic

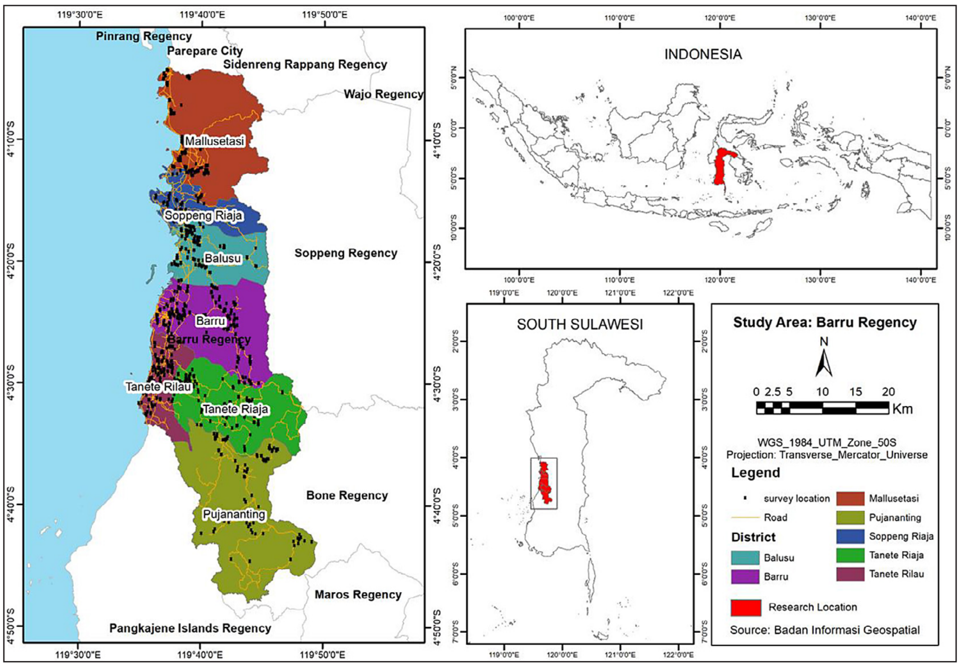

Figure 1. Location of study area in Barru Regency for rice land suitability analysis

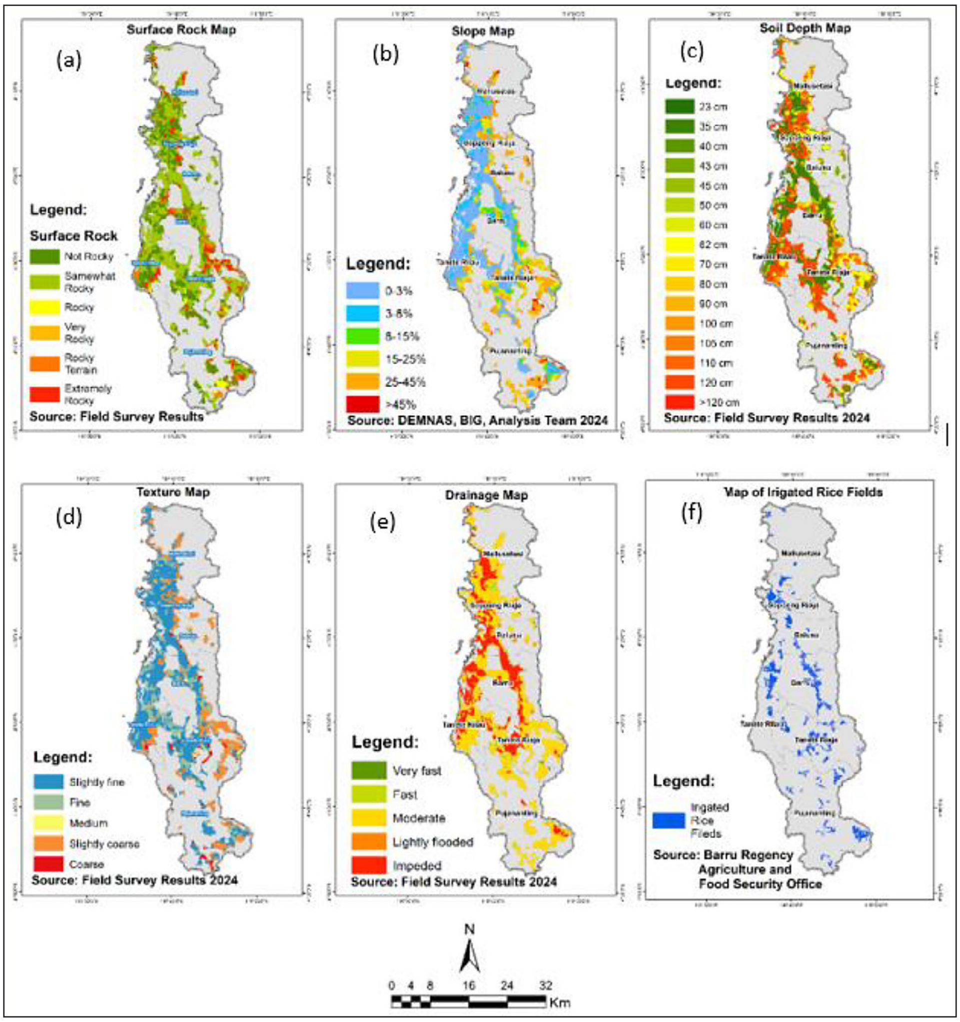

Figure 2. Land characteristics maps: (a) surface rock map, (b) slope map, (c) soil depth map, (d) soil texture map, (e) drainage map, and (f) irrigated rice field map. Data sources for each map are provided in Table 1

carbon, bulk density, potassium, nitrogen, phosphorus, and pH) were not collected directly from these 530 sampling points. Instead, they were obtained from soil surveys conducted by the Barru Regency Agriculture and Food Security Office and validated through laboratory analyses at the Soil Laboratory of Hasanuddin University (2024). These parameters were then spatially integrated into GIS framework at a 30 × 30 m raster resolution. The spatial distribution of the sampling points is visualized in the inset map in Figure 1, illustrating their coverage across the study area.

Data

A comprehensive dataset was compiled from various sources, as summarized in Table 1. Soil parameters were obtained from the Barru Regency Agriculture and Food Security Office and validated through field surveys represented by one sampling point per district and laboratory analysis conducted at the Soil Laboratory of Hasanuddin University (2024). Slope data were extracted from the digital elevation model (DEM) provided by the Regional Development Planning Agency (2023). Rainfall data (2015–2024) were processed from CHIRPS using GIS tools. Agricultural production and irrigation data were collected from the Barru Regency Agriculture Office and Statistics Indonesia (BPS) (2019–2024).

Technical irrigation, defined as rice fields managed by engineered systems (e.g., canals and sprinklers), was represented as a binary variable (irrigated = 1, non-irrigated = 0) based on the 2024 irrigation coverage map and direct field surveys conducted at 530 sampling points. To address spatial misalignment between productivity data and soil parameter data, all datasets were spatially standardized to a common resolution of 30 × 30 m using GIS tools. Vector-based data such as administrative boundaries, irrigation infrastructure, and land cover were converted to raster format using the nearest-neighbor or majority resampling methods, depending on data type, to preserve categorical integrity. Raster alignment was based on a common spatial reference and grid origin to ensure that each pixel represented the same ground location across all layers. This

Table 1. Types and sources of data

Type of data	Description	Source
Administrative boundaries	Sub-district and village boundaries	Regional Development Planning Agency (2023)
Land characteristics and soil fertility	Slope, drainage, texture, soil depth, surface rock, cec, c organic, bulk density, potassium, nitrogen, phosphorus, pH	Soil surveys and laboratory analysis (Soil laboratory of the Faculty of Agriculture, Hasanuddin University)
DEMNAS	DEMNAS was utilized to generate the slope map	https://tanahair.indonesia.go.id/portal-web/
Rainfall map	Average annual rainfall (2015–2024)	https://www.chc.ucsb.edu/data/chirps Processed from CHIRPS using GIS
RTRW map	Designated food production areas	Regional Development Planning Agency (2023)
Agricultural production	Production data for LQ and trend analysis	Barru Regency Agriculture Office and BPS (2019–2023)
Rice productivity (ton/ha)	Ground validation using mobile mapping: https://arcg.is/1yPKfH1	Field observation
Irrigation area map	Rice field areas supported by technical irrigation systems	Barru Regency Agriculture Office and BPS (2024)

Table 2. Soil fertility distribution

Classification	Soil fertility (ha)					
	CEC	C organic	Bulk density	Potassium	Nitorgen	Posfor
Very low	0.00	2872.75	0.00	0.00	62.90	13415.52
Low	15563.63	28879.23	11124.87	37260.59	42478.19	17577.08
Moderate	31126.34	15583.62	23196.97	10075.01	4794.51	16343.00
Fairly high	0.00	0.00	12523.41	0.00	0.00	0.00
High	645.63	0.00	447.36	0.00	0.00	0.00
Very high	0.00	0.00	42.99	0.00	0.00	0.00

preprocessing step minimized potential integration bias during the overlay and analysis stages.

Table 1 summarizes the types and sources of data used in this study, ranging from administrative boundaries and soil properties to rainfall distribution and designated food production areas. Agricultural production data and rice yield estimates were obtained from historical records and ground-truth surveys. In addition, an irrigation area map was developed using updated information from local agricultural offices, reflecting the extent and coverage of technical irrigation systems as of 2024. Collectively, this dataset forms the basis for subsequent spatial modeling and predictive analysis using machine-learning approaches.

Mapping units

Following the FAO guidelines, mapping units were defined based on 12 soil parameters extracted from the soil samples and the DEM. These parameters were processed into raster layers with a resolution of 30 × 30 m using GIS, resulting in 245 distinct mapping units across the study area. The spatial distribution of land characteristics and soil fertility parameters are presented in Figures 2 and 3, each classified based on the standard FAO evaluation criteria.

Land suitability assessment

Land suitability and rice productivity were assessed using a RF model within the GIS framework. The RF model was configured with 100 decision trees (n_estimators = 100), a maximum depth of 10 (max_depth = 10), and square-root feature selection (max_features = sqrt) to mitigate overfitting. A Multi-Output Classifier was employed to simultaneously predict FAO-based land suitability (S1 = 1, S2 = 2, S3 = 3, N = 4) and rice productivity (1 = 1.0–2.5; 2 = 2.6–4.5; 3 = 4.6–7.0; 4 = 7.1–10 tons/ha). The dataset was normalized using Standard Scaler to ensure feature comparability. The model was trained on 530 sampling points, and technical irrigation was included as an additional predictor to assess managerial impact. A 5-fold cross-validation was used to evaluate the performance, complemented by out-of-bag (OOB) scores to assess generalization. Permutation Importance (10 repetitions) was applied to quantify each predictor's mean decrease in accuracy (MDA) and to identify key drivers. The modeling workflow depicted in Figure 4 includes spatial data collection, GIS preprocessing, data partitioning, RF training, validation, and predictive mapping.

Although RF is inherently scale-invariant due to its decision tree-based structure, which splits

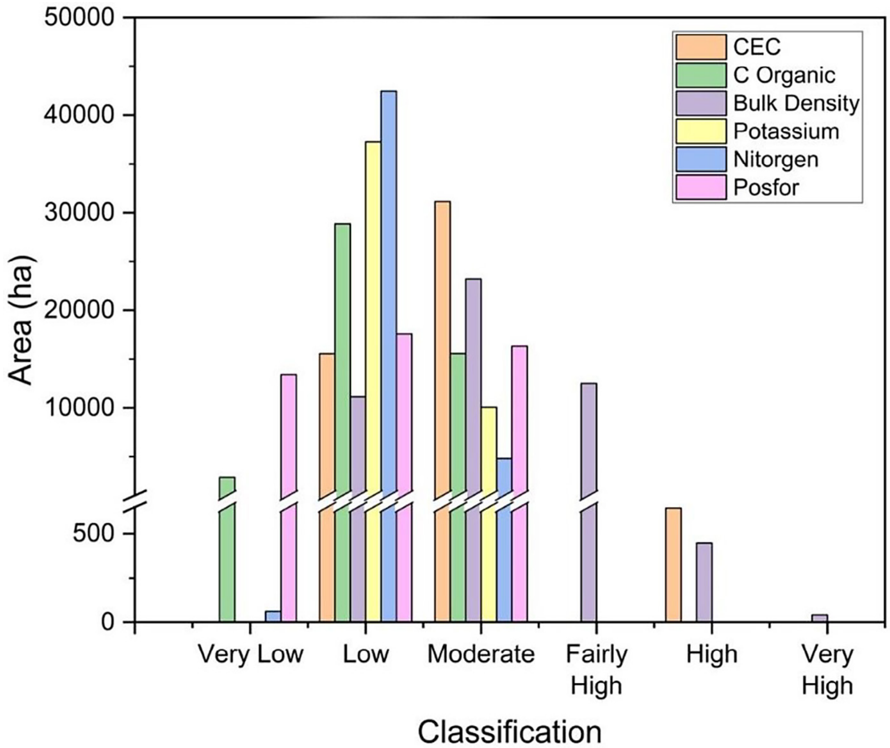

Figure 3. Distribution area for each soil fertility parameter

features based on relative thresholds rather than absolute values, using Standard Scaler in this study's preprocessing workflow is justified for several reasons. First, standardizing the 12 GIS-derived parameters and technical irrigation data ensures numerical stability and consistency across heterogeneous features (e.g., soil pH, cation exchange capacity, and irrigation metrics) that have different units and ranges. Placing all predictors on a standard scale facilitates a more straightforward interpretation and comparison of feature importance metrics, such as MDA from Permutation Importance. Second, standardization mitigates potential biases in the Multi-Output Classifier's handling of simultaneous land suitability and rice productivity predictions, where unscaled features could disproportionately influence the model's internal computations, particularly during cross-validation. Finally, preprocessing with Standard Scaler aligns with best practices for integrating RF within a GIS framework, where normalized inputs enhance compatibility with spatial data processing and visualization tools, improving the robustness of the predictive mapping workflow depicted in Figure 4.

Land suitability mapping

The workflow for predicting FAO-based land suitability and rice productivity in Barru Regency using the RF model is illustrated in Figure 4, encompassing the full research process from data collection and geospatial preprocessing to machine learning development and predictive mapping. This integrated approach enhances precision agriculture in tropical environments, where resource management must account for diverse climatic and geographic conditions. The methodology includes data preprocessing, defining observational parameters, rigorous model validation, and ensuring robust predictions aligned with sustainability and food security goals (Baroudy et al., 2020; Ramadhani et al., 2021; Taghizadeh-Mehrjardi et al., 2020). Leveraging RF's capacity

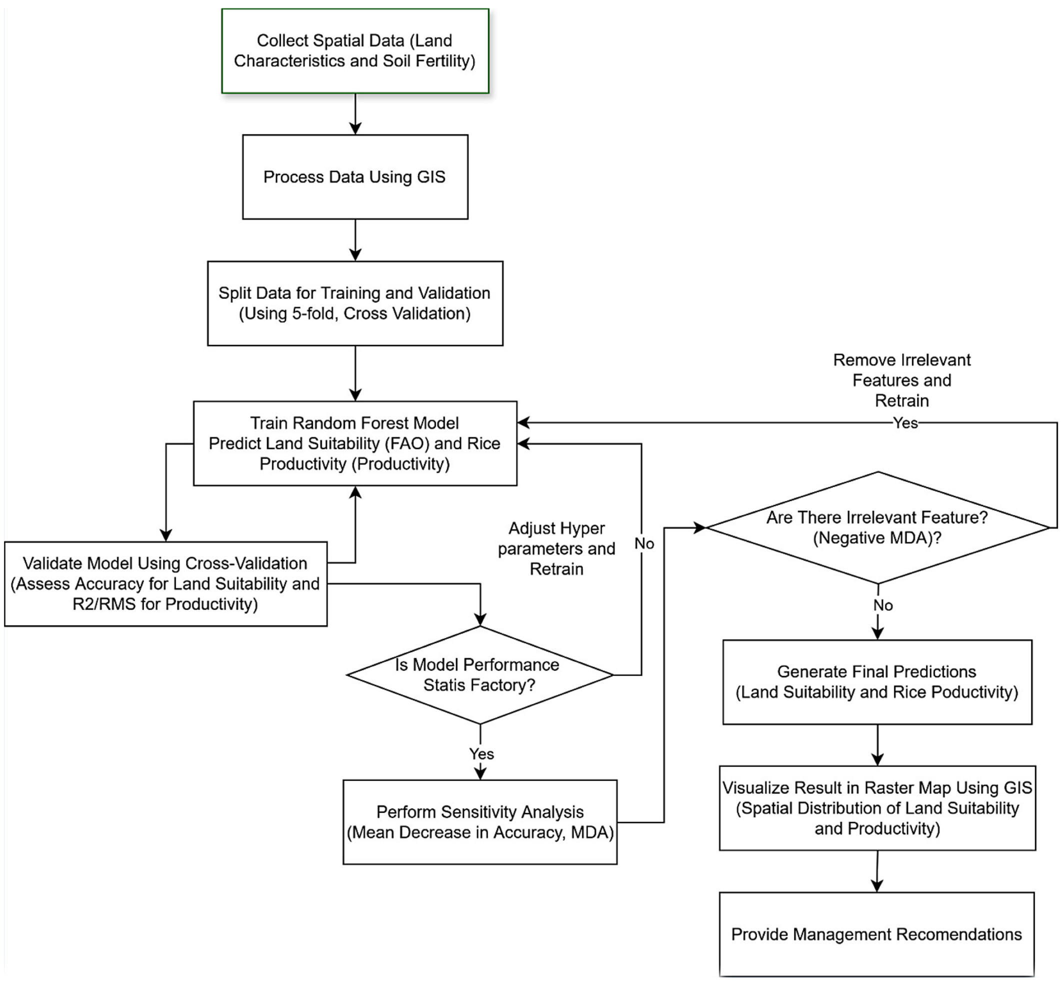

Figure 4. Flowchart of the research process

to process complex, multidimensional datasets, this study incorporates FAO land suitability criteria such as soil type, nutrient retention, and hydrotopography to inform land management strategies. This evidence-based framework provides stakeholders with actionable insights into promoting sustainable farming, increasing productivity, and adapting to future climate variability in the Barru Regency (Kumar et al., 2023; Makungwe et al., 2021).

Comprehensive spatial data collection was conducted to generate vector and raster datasets, as detailed in Section 2.2, and the results are summarized in Table 1. This included 11 vector layers capturing land characteristics, such as drainage, texture, and soil depth, as well as fertility indicators, such as CEC, organic carbon, and pH. To ensure consistency with the slope dataset, originally at 10 × 10 m and resampled to 30 × 30 m, all vector layers were converted to the raster format, reducing the computational load while maintaining spatial accuracy (Wadoux et al., 2020). Standardizing data formats ensures a uniform spatial extent and alignment, enabling seamless integration within a GIS environment (Adekiya et al., 2022). From these harmonized layers, 12 key soil parameters were extracted at 530 field survey points, providing essential ground-truth data for training and validating the Random Forest model. This rigorous preprocessing approach highlights the critical role of data standardization in producing reliable and accurate spatial modeling outcomes (Adugna et al., 2022; Kumar et al., 2023).

Following data preparation, the dataset was split using 5-fold cross-validation to ensure robust model evaluation and reduce overfitting, followed by the integration of spatial cross-validation techniques to prevent data leakage in spatially autocorrelated environments. The Random Forest (RF) model, configured with 100 decision trees (n_estimators = 100), a maximum depth of 10, and square-root-based feature selection (max_features = √), was implemented within a multi-output classifier to simultaneously predict FAO land suitability classes (S1, S2, S3, N) and rice productivity categories (1–2.5, 2.6–4.5, 4.6–7.0, 7.1–10 tons/ha), thereby addressing both land evaluation and yield forecasting (Baroudy et al., 2020; Das, 2024). Feature importance was assessed using Permutation Importance with 10 repetitions to identify key predictors influencing model performance (Ahsan et al., 2021). The trained model was then applied across the study area to generate predictive raster maps within a GIS environment, enabling spatial interpretation of suitability and productivity. This approach provides valuable insights for informed land management and strategic agricultural planning in Barru Regency, aligned with best land-use evaluation practices (Handoko et al., 2024; Lukman et al., 2021; Ramadhani et al., 2021; Suntoro et al., 2023).

Accuracy assessment

Model performance was evaluated using multiple metrics: overall accuracy, precision, recall, and F1-score for land suitability and root mean square error (RMSE) and coefficient of determination (R^2) for productivity. The RF configuration was tested across variations in n_estimators and max_depth to ensure stability. OOB scores and 5-fold cross-validation provided the mean accuracy and standard deviation to address overfitting, ensuring robust assessment of predictive capabilities.

RESULTS AND DISCUSSION

Results

The random forest (RF) model was evaluated for its predictive performance in assessing FAO-based land suitability and rice productivity in Barru Regency, using 12 GIS-derived soil parameters: slope, drainage, texture, soil depth, surface rock, CEC, organic carbon, bulk density, potassium, nitrogen, phosphorus, and pH. Based on 5-fold cross-validation, RF achieved an accuracy of 0.95 ± 0.01 for FAO land suitability and 0.32 ± 0.04 for rice productivity, demonstrating high stability for the former due to the consistent performance across folds (variation of ± 0.01). The OOB scores for RF were 0.96 FAO land suitability and 0.33 productivity, which closely aligned with the cross-validation results, suggesting minimal overfitting for suitability predictions. However, productivity predictions exhibited greater variability (± 0.04), potentially owing to the spatial heterogeneity in the dataset.

Sensitivity analysis using Permutation Importance (10 repetitions) revealed the distinct contributions of the 12 soil parameters to each target. For FAO land suitability, RF identified CEC (0.0123 ± 0.0134), slope (0.0094 ± 0.0112), soil depth (0.0028 ± 0.0043), and bulk density (0.0009 ± 0.0028) as the most influential parameters, reflecting their critical roles in determining land suitability under the FAO framework. Other parameters, such as drainage, texture, and surface rock, exhibited negligible impacts (MDA < 0.001), suggesting limited discriminatory power in this context. For rice productivity, the RF-highlighted slope (0.0613 ± 0.0330), potassium (0.0113 ± 0.0102), CEC (0.0104 ± 0.0181), and drainage (0.0009 ± 0.0200) were the primary contributors. However, several parameters, including soil depth, organic carbon, and bulk density, yielded negative MDA values (e.g., soil depth: -0.0368 ± 0.0157), indicating that these features introduced noise and potentially contributed to overfitting in productivity predictions. The negative MDA values for soil depth, organic carbon, and bulk density in the RF model for rice productivity indicate that these features introduced noise and contributed to overfitting. They were most likely dropped from the final model to improve predictive performance and generalization, focusing on the top contributors like slope, potassium, CEC, and drainage. Some investigation into the cause of the negative MDA values (e.g., data quality, redundancy) was likely performed to confirm this decision, but the primary action was to exclude these features. For future work, revisiting data quality, exploring feature interactions, and testing alternative models could help address these issues and potentially reintroduce these features if their data quality improves.

The confusion matrix shows the number of correct and incorrect predictions for each class, while the F1-score is the harmonic mean of precision and recall, providing a balanced overview of the model's performance for each class.

- F1-score per Class:
- Class 1 (S1): 0.39
- Class 2 (S2): 0.49
- Class 3 (S3): 0.12
- Class 4 (N): 0.15

Confusion Matrix:

$$\begin{bmatrix} 26 & 24 & 5 & 8 \\ 33 & 47 & 13 & 11 \\ 8 & 12 & 3 & 2 \\ 5 & 5 & 2 & 3 \end{bmatrix}$$

- S2 has the highest F1-score (0.49), indicating the best performance among all classes. This is supported by the relatively high number of correct predictions (47) compared to other classes, although there is still significant confusion with class S1 (33 misclassifications).
- S1 has an F1-score of 0.39, which is lower than S2. The model often confuses S1 with S2 (24 misclassifications), possibly due to feature similarities between the two classes.
- S3 and N have very low F1-scores (0.12 and 0.15), indicating poor performance. This is reflected in the very small number of correct predictions (3 for both classes) and the high number of misclassifications (22 for S3 and 12 for N).

Technical irrigation parameters were incorporated into the RF model to explore the potential of managerial factors to improve productivity predictions. This adjustment resulted in a 6% increase in cross-validation accuracy for productivity, from 0.32 ± 0.04 to 0.38 ± 0.06, highlighting the importance of irrigation management as a determinant of rice productivity in constrained tropical lands. The discussion on technical irrigation parameters is integrated to assess the suitability of rice productivity predictions in relation to agricultural crop management. These parameters are sourced from data provided by the Barru Regency Agriculture Office (2024) and analyzed using the random forest model, which demonstrates a 6% increase in accuracy (from 0.32 to 0.38 ± 0.06). The impact on yields, particularly in low-productivity areas such as Balusu and Tanete Rilau, was shown in Table 4. The spatial distribution of land suitability classes, derived from RF predictions, revealed that S3 class dominated the study area, covering 34,836.62 ha, followed by classes N (8,303.68 ha) and S2 (2,722.62 ha). Overlay analysis with sub-district administrative boundaries showed that the Barru Sub-district had the largest area of S2-class land among the seven sub-districts, spanning 1,181.16 ha. In contrast, the other sub-districts had S2 areas below 500 ha (Table 3). For productivity, the majority of the area fell into the 2.6–4.5 tons/ha class (25,081.02 ha), with Barru Sub-district exhibiting the largest extent of high-productivity land (7.1–10 tons/ha: 2,346.88 ha) (Table 4). Predictive maps generated from RF further corroborated these findings, producing sharp boundaries for FAO suitability classes and clear delineations of the productivity distributions (Figure 5).

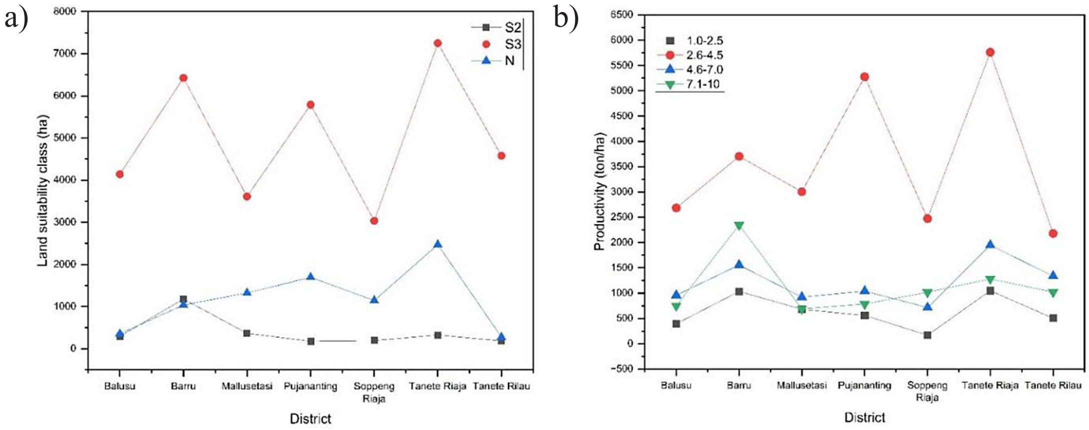

Figure 5. The distribution of (a) land suitability classes and (b) productivity classes using random forest

Table 3. Land suitability class for each sub-district

District	Land suitability class (ha)		
	S2	S3	N
Balusu	293.98	4,140.55	353.45
Barru	1,181.16	6,424.98	1,038.37
Mallusetasi	364.95	3,610.21	1,323.65
Pujananting	177.32	5,789.72	1,694.85
Soppeng Riaja	194.25	3,035.07	1,145.59
Tanete Riaja	324.49	7,250.12	2,470.04
Tanete Rilau	185.92	4,580.23	276.60
Grand Total	2,722.62	34,836.62	8,303.69

Table 4. Productivity in each sub-district

District	Area Productivity (ha)			
	1.0–2.5 (ton/ha)	2.6–4.5 (ton/ha)	4.6–7.0 (ton/ha)	7.1–10 (ton/ha)
Balusu	393.94	2,682.22	959.31	749.64
Barru	1,035.89	3,702.52	1,556.75	2,346.88
Mallusetasi	681.17	3,003.65	922.24	689.13
Pujananting	559.31	5,275.94	1,040.24	785.85
Soppeng Riaja	169.26	2,472.28	714.92	1,017.28
Tanete Riaja	1,048.60	5,762.02	1,948.17	1,283.07
Tanete Rilau	505.10	2,179.09	1,336.20	1,019.95
Grand Total	4,393.31	25,081.02	8,479.91	7,893.33

Discussion

The exceptional performance of RF in predicting FAO-based land suitability highlights its remarkable ability to discern complex patterns within GIS data, a capability corroborated by global studies such as Baroudy et al. (2020) and Taghizadeh-Mehrjardi et al. (2020), which affirm RF's effectiveness in geospatial classification tasks for agriculture. The model's heightened sensitivity to key environmental factors like CEC, slope, and soil depth aligns with FAO guidelines and regional findings by Adekiya et al. (2022) and Makungwe et al. (2021), emphasizing the critical role of topography and soil fertility in land management within resource-constrained regions like sub-Saharan Africa, where these factors help address challenges such as soil erosion and nutrient

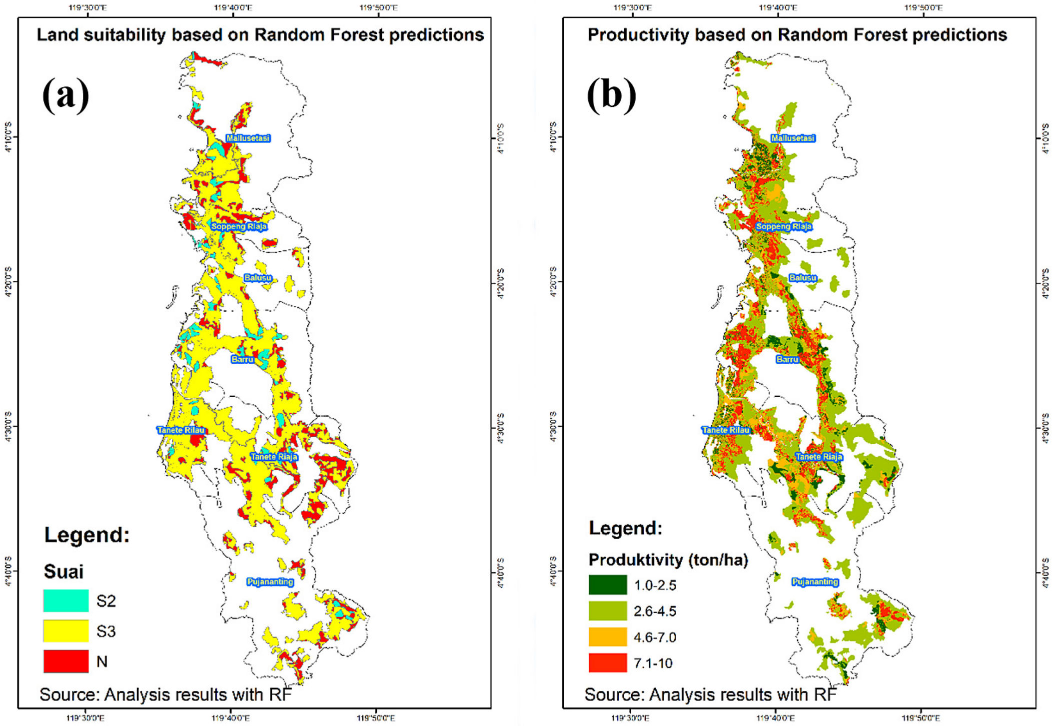

Figure 6. Distribution map of: (a) land suitability classes, and (b) productivity classes using random forest

depletion, thereby supporting sustainable agronomic decision-making and enhancing food security.

Moreover, the substantial influence of slope (Mean Decrease Accuracy, MDA: 0.0094 ± 0.0112) elucidates its critical role in assessing land viability for rice cultivation, notably in regions characterized by pronounced topographical variations, such as the Barru Regency. Here, steep slopes pose irrigation challenges and increase the soil erosion risk (Handoko et al., 2024). Similarly, CEC's strong contribution of CEC (MDA: 0.0123 ± 0.0134) signals its essential nutrient retention function, particularly in tropical environments that are susceptible to high leaching (Dalle et al., 2021). The graphical representation of the soil fertility parameters further corroborates these insights, reinforcing that effective land management strategies must consider these variables (Dalle et al., 2021; Taghizadeh-Mehrjardi et al. 2020). Conversely, the negligible effects of parameters such as drainage and texture (MDA < 0.001) suggest limited variability or relevance in this specific context, warranting additional scrutiny of their spatial distribution and measurement methodologies.

The comparatively lower accuracy of the RF model in predicting rice productivity (0.32 ± 0.04) reflects inherent challenges associated with complex and potentially imbalanced datasets. These challenges are further highlighted by the negative MDA values for crucial parameters, such as soil depth (-0.0368 ± 0.0157). Previous studies indicate that negative MDA values signal the presence of noise in the dataset, often stemming from data inconsistencies or unaccounted spatial variability during GIS preprocessing (Baroudy et al., 2020). Inaccuracies in soil depth measurements could arise from poorly conducted field surveys or lack relevance in regions characterized by uniform soil profiles, which may lead to overfitting of the RF model to these misleading data patterns. Conversely, the comparatively high MDA values associated with slope (0.0613 ± 0.0330) and CEC (0.0123 ± 0.0134) suggest that these factors play a pivotal role in influencing rice productivity. This assertion was supported by the findings of Baroudy et al. Baroudy et al. (2020) emphasized the significant impact of topographical features and soil nutrient availability on agricultural yields. Furthermore, integrating comprehensive land suitability assessments is essential for optimizing agricultural productivity. Such assessments involve multi-criteria evaluations, which elucidate the intricate relationship between soil characteristics and agricultural output (Veisi et al., 2024). The complexity embedded within spatial data models necessitates robust machine learning algorithms that can handle the variability and depth of agricultural datasets, as evidenced by various model comparisons in the literature (Makungwe et al., 2021). Therefore, a nuanced understanding and careful consideration of parameter weighting in predictive models are fundamental for enhancing the accuracy of crop productivity forecasts.

The integration of technical irrigation parameters led to a 6% improvement in rice productivity prediction accuracy (from 0.32 to 0.38 ± 0.06), emphasizing the importance of managerial inputs in overcoming natural constraints in tropical agriculture. This suggests that irrigation management can counterbalance limiting factors such as variable soil depth by ensuring a steady water supply for rice cultivation. Spatial analysis revealed that the S3 land suitability class dominated the Barru Regency (34,836.62 ha), as shown in 4 and Figures 5(a) and 6(a), indicating moderate constraints, mainly due to steep slopes and low soil fertility, corroborated by the high MDA values for slope and CEC. Targeted interventions, such as terracing and soil amendments, are especially needed in sub-districts such as Tanete Riaja (7,250.12 ha) and Pujananting (5,789.72 ha), which have the largest S3 areas. Conversely, Barru Sub-district shows a higher share of S2 land (1,181.16 ha) and high rice productivity (7.1–10 tons/ha: 2,346.88 ha), likely due to favorable terrain and better irrigation access. Meanwhile, the 2.6–4.5 tons/ha class dominates rice productivity (25,081.02 ha), indicating moderate yield potential, while sub-districts such as Balusu and Tanete Rilau show significantly low-productivity zones (1.0–2.5 tons/ha), likely tied to poor soil and irrigation access (Table 4, Figures 5(b) and 6(b)). These findings underscore the need for spatially targeted policy actions, including expanding irrigation infrastructure in low-productivity areas to enhance yields and prioritizing land management interventions in S3-class zones to optimize agricultural performance across the Barru Regency.

Integrating GIS and machine learning, particularly RF, has significantly advanced precision agriculture through effective land suitability mapping, yet challenges like low productivity prediction accuracy underscore the need for improved data inputs, addressing class imbalances, and incorporating critical climatic variables such as rainfall seasonality to enhance model performance (Taghizadeh-Mehrjardi et al., 2020), while also highlighting the importance of robust preprocessing like data scaling and enrichment to mitigate inaccuracies; however, limitations such as the reliance on specific datasets with limited spatial or temporal coverage, the absence of comprehensive climatic data, and the lack of extensive comparison with other algorithms may restrict generalizability, underestimate productivity risks, and hinder scalability, necessitating future research to adopt comprehensive frameworks combining diverse techniques like RF and support vector machines (SVM) with rigorous data refinement to address spatial heterogeneity and climatic influences, ultimately unlocking the full potential of GIS-based machine learning for sustainable productivity gains (Fei et al., 2023; Kok et al., 2021).

CONCLUSIONS

This study advances GIS-machine learning (GIS-ML) applications in precision agriculture by showcasing the RF algorithm's strong performance in predicting FAO-based land suitability in Barru Regency, where it achieved an impressive cross-validation accuracy of 0.95 ± 0.01, effectively leveraging GIS data to produce reliable suitability maps that support strategic interventions like prioritizing irrigation in low-productivity sub-districts (e.g., Balusu and Tanete Rilau) and improving soil conditions in S3-dominated areas; however, its performance in rice productivity prediction was notably limited, with a lower accuracy of 0.32 ± 0.04, which improved to 0.38 ± 0.06 upon integrating technical irrigation parameters, highlighting the challenges of noisy, complex datasets, class imbalances, spatial heterogeneity, and missing climatic variables like rainfall seasonality, indicating that while RF excels in suitability classification, its struggles with multi-target noisy data necessitate advanced approaches such as oversampling, high-resolution climatic data inclusion, and exploration of deep learning models like convolutional neural networks (CNNs) to enhance productivity predictions and strengthen food security in tropical, infrastructure-limited regions.

Acknowledgements

The authors thank the Department of Agriculture and Food Security of Barru Regency for providing the necessary data and support for this study. We also thank the farmers and local communities in the Barru Regency for their cooperation during the field surveys. Additionally, we acknowledge the support of Universitas Hasanuddin for providing research facilities and resources and facilitating the Kedaireka research program. This study was funded by the Program Dana Padanan 2024 Kedaireka from the Directorate General of Higher Education, Research, and Technology, under contract number 1936/UN4.11/HK.07.00/2024, through a partnership between Universitas Hasanuddin and the Department of Agriculture and Food Security of Barru Regency, and we are grateful for their financial support. Finally, we appreciate the valuable feedback from colleagues and reviewers, which helped improve the quality of this manuscript.

REFERENCES

1. Adekiya, A. O., Adegbite, K. A., Aremu, C., Adebiyi, O. T., Alori, E. T., Ayorinde, B. B., Olaniyan, F., Makun, B., Ajayi, G. (2022). Land Use capability and suitability classifications of a plinthic alfisol of a derived savannah ecology of Nigeria. *The Open Agriculture Journal, 16*(1). https://doi.org/10.2174/18743315-v16-e2204220
2. Adugna, T., Xu, W., Fan, J. (2022). Comparison of random forest and support vector machine classifiers for regional land cover mapping using coarse resolution FY-3C images. *Remote Sensing, 14*(3), 574. https://doi.org/10.3390/rs14030574
3. Baroudy, A. A. El, Ali, Abdelraouf. M., Mohamed, E. S., Moghanm, F. S., Shokr, M. S., Savin, I., Poddubsky, A., Ding, Z., Kheir, A. M. S., Aldosari, A. A., Elfadaly, A., Dokukin, P., Lasaponara, R. (2020). Modeling land suitability for rice crop using remote sensing and soil quality indicators: The case study of the Nile Delta. *Sustainability, 12*(22), 9653. https://doi.org/10.3390/su12229653
4. Dalle, J., Hastuti, D., Akmal, F. (2021). Evaluation of peatland suitability for rice cultivation using matching method. *Polish Journal of Environmental Studies, 30*(3), 2041–2047. https://doi.org/10.15244/pjoes/127420
5. Das, B. (2024). An assessment on land suitability for rice cultivation using analytical hierarchy process in the Sivasagar District of Assam, India. *International Research Journal of Multidisciplinary Scope, 5*(4), 942–959. https://doi.org/10.47857/irjms.2024.v05i04.01517
6. David, R. M., Rosser, N. J., Donoghue, D. N. M. (2022). Remote sensing for monitoring tropical dryland forests: A review of current research, knowledge gaps and future directions for Southern Africa. *Environmental Research Communications, 4*(4). https://doi.org/10.1088/2515-7620/ac5b84
7. Fei, S., Hassan, M. A., Xiao, Y., Su, X., Chen, Z., Cheng, Q., Duan, F., Chen, R., Ma, Y. (2023). UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat. *Precision Agriculture, 24*(1), 187–212. https://doi.org/10.1007/s11119-022-09938-8
8. Gbode, I. E., Ajayi, V. O., Coulibaly, A., Abel, D., Ziegler, K., Weber, T., Traore, S. B., Raji, I. A., Paeth, H. (2025). Analysis of crop suitability index for current and future climates using statistically downscaled <scp>CMIP6</scp> outputs over Africa. *Meteorological Applications, 32*(1). https://doi.org/10.1002/met.70022
9. Handoko, D., Istiarto, Karlina. (2024). Land suitability studies for paddy cultivation in Dadahup Swampy Irrigation Area. *IOP Conference Series: Earth and Environmental Science, 1311*(1), 012019. https://doi.org/10.1088/1755-1315/1311/1/012019
10. Khan, T., Tian, W., Zhou, G., Ilager, S., Gong, M., Buyya, R. (2022). Machine learning (ML)-centric resource management in cloud computing: a review and future directions. *Journal of Network and Computer Applications, 204*(April). https://doi.org/10.1016/j.jnca.2022.103405
11. Kok, Z. H., Mohamed Shariff, A. R., Alfatni, M. S. M., Khairunniza-Bejo, S. (2021). Support Vector Machine in Precision Agriculture: a Review. *Computers and Electronics in Agriculture, 191*, 106546. https://doi.org/10.1016/j.compag.2021.106546
12. Kumar, V., Kedam, N., Sharma, K. V., Mehta, D. J., Caloiero, T. (2023). Advanced machine learning techniques to improve hydrological prediction: a comparative analysis of streamflow prediction models. *Water, 15*(14), 2572. https://doi.org/10.3390/w15142572
13. Limpo, S. Y., Fahmid, I. M., Fattah, A., Rauf, A. W., Surmaini, E., Muslimin, Saptana, Syahbuddin, H., Andri, K. B. (2022). Integrating indigenous and scientific knowledge for decision making of rice farming in South Sulawesi, Indonesia. *Sustainability (Switzerland), 14*(5). https://doi.org/10.3390/su14052952
14. Lukman, K. M., Uchiyama, Y., Quevedo, J. M. D., Harding, D., Kohsaka, R. (2021). Land use changes assessment using a triangulated framework: perception interviews, land-use/land cover observation, and spatial planning analysis in Tanjung Batu and Derawan Island, Indonesia. *Human Ecology, 2021*, 551–564. https://doi.org/10.1007/s10745-021-00253-w
15. Makungwe, M., Chabala, L. M., Van Dijk, M., Chishala, B. H., Lark, R. M. (2021). Assessing land suitability for rainfed paddy rice production in Zambia. *Geoderma Regional, 27*, e00438. https://doi.

org/10.1016/j.geodrs.2021.e00438

16. McDowell, R. W., Snelder, T., Harris, S., Lilburne, L., Larned, S. T., Scarsbrook, M., Curtis, A., Holgate, B., Phillips, J., Taylor, K. (2018). The land use suitability concept: Introduction and an application of the concept to inform sustainable productivity within environmental constraints. *Ecological Indicators, 91*(February), 212–219. https://doi.org/10.1016/j.ecolind.2018.03.067
17. Otsuka, K. (2021). Strategy for transforming indonesian agriculture. *Bulletin of Indonesian Economic Studies, 57*(3), 321–341. https://doi.org/10.1080/00074918.2021.2002387
18. Pichler, M., Boreux, V., Klein, A., Schleuning, M., Hartig, F. (2019). Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. 11(2), 281–293. https://doi.org/10.1111/2041-210x.13329
19. Ramadhani, E. E., Sujono, J., Taryono. (2021). Agriculture land suitability of tidal swampy area at Palingkau Irrigation Area in Central Kalimantan Province for national food estate program. *IOP Conference Series: Earth and Environmental Science, 930*(1), 012069. https://doi.org/10.1088/1755-1315/930/1/012069
20. Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., Khan, N. (2021). A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. *IEEE Access, 9*, 63406–63439. https://doi.org/10.1109/ACCESS.2021.3075159
21. Shahhosseini, M., Hu, G., Huber, I., Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US corn belt. *Scientific Reports, 11*(1), 1606. https://doi.org/10.1038/s41598-020-80820-1
22. Shoaib, M. R., Emara, H. M., Zhao, J. (2023). Revolutionizing global food security: empowering resilience through integrated ai foundation models and data-driven solutions. *ArXiv*, 1–32. https://app.dimensions.ai/details/publication/pub.1165449282
23. Suntoro, S., Dewi, W. S., Syamsiyah, J., Mujiyo, M., Herdiansyah, G., Widijanto, H., Herawati, A. (2023). Evaluation of land suitability as the basis for sustainable crop development. *IOP Conference Series: Earth and Environmental Science, 1228*(1), 012021. https://doi.org/10.1088/1755-1315/1228/1/012021
24. Taghizadeh-Mehrjardi, R., Nabiollahi, K., Rasoli, L., Kerry, R., Scholten, T. (2020). Land suitability assessment and agricultural production sustainability using machine learning models. *Agronomy, 10*(4), 573. https://doi.org/10.3390/agronomy10040573
25. Taghizadeh-Mehrjardi, R., Schmidt, K., Amirian-Chakan, A., Rentschler, T., Zeraatpisheh, M., Sarmadian, F., Valavi, R., Davatgar, N., Behrens, T., Scholten, T. (2020). Improving the spatial prediction of soil organic carbon content in two contrasting climatic regions by stacking machine learning models and rescanning covariate space. *Remote Sensing, 12*(7), 1095. https://doi.org/10.3390/rs12071095
26. Tziolas, N., Tsakiridis, N., Chabrillat, S., Demattê, J. A. M., Ben-Dor, E., Gholizadeh, A., Zalidis, G., van Wesemael, B. (2021). Earth observation data-driven cropland soil monitoring: a review. *Remote Sensing, 13*(21). https://doi.org/10.3390/rs13214439
27. Veisi, A., Khoshbakht, K., Veisi, H., Talarposhti, R. M., Tanha, R. H. (2024). Integrating farmers' and experts' perspectives for soil health-informed decision-making in conservation agriculture systems. *Environment Systems and Decisions, 44*(2), 199–214. https://doi.org/10.1007/s10669-023-09923-0
28. Wadoux, A. M. J.-C., Minasny, B., McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews, 210*, 103359. https://doi.org/10.1016/j.earscirev.2020.103359
29. Zhang, J., Yong, H., Lv, N. (2024). Balancing productivity and sustainability: insights into cultivated land use efficiency in Arid Region of Northwest China. *Journal of the Knowledge Economy, 15*(3), 13828–13856. https://doi.org/10.1007/s13132-023-01652-8