Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Application of Machine Learning Random Forest Algorithm in Digital Soil Mapping of Key Soil Properties for Sustainable Tea Productivity

Finias F. Mwesige

finiasfidelis@gmail.com

Sokoine University of Agriculture

Boniface H. J. Massawe

Sokoine University of Agriculture

Hilda G. Sanga

Sokoine University of Agriculture

Braison E. Mjanja

Tea Research, Tanzania Agricultural Research Institute

Erasto Focus

Mwalimu Nyerere University of Agriculture and Technology

Research Article

Keywords: Digital soil mapping, Random Forest, SCORPAN, Soil properties, Tea cultivation, GIS

Posted Date: September 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10639002/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Application of Machine Learning Random Forest Algorithm in Digital Soil Mapping of Key Soil Properties for Sustainable Tea Productivity

Finias F. Mwesige[1,2*]; Boniface H. J. Massawe[1]; Hilda G. Sanga[1]; Braison E. Mjanja[3]; Erasto Focus[4]

[1]Department of Soil and Geological Sciences, Sokoine University of Agriculture, P.O. Box 3008 CHUO KIKUU, Morogoro, Tanzania

[2]Department of Agricultural Sciences, Mwalimu Nyerere University of Agriculture and Technology,

P.O Box 976, Musoma, Tanzania

[3] Tea Research, Tanzania Agricultural Research Institute, P. O Box 1571, Dodoma, Tanzania

[4]Department of Physics, Mwalimu Nyerere University of Agriculture and Technology, P.O Box 976, Musoma, Tanzania

---

ABSTRACT

Digital soil mapping (DSM) plays a crucial role in optimizing agricultural resource use including sustainable tea productivity. This study predicted key soil properties using a Random Forest (RF) machine-learning algorithm within a DSM framework, integrating soil, climate, organisms, relief, parent material, age, and spatial position (SCORPAN)-based environmental covariates derived from digital elevation model (DEM)-based topographic variables, Sentinel-2A spectral indices, and spatial soil information. A total of 64 topsoil (0–30 cm) samples collected during a soil survey in Ganyange Ward, Tarime District, Tanzania,

were analysed in the laboratory. Predicted soil properties included soil reaction (pH), soil organic carbon (SOC), total nitrogen (N), available phosphorus (P), exchangeable base cations (Ca, Mg, and K), cation exchange capacity (CEC), and soil texture fractions (sand, silt, and clay). Model performance was evaluated using spatial cross-validation, where approximately 75% of samples were used for training and 25% for validation in each fold. Accuracy metrics included root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). The RF model demonstrated moderate to strong predictive performance for some soil properties, with positive R^2 values ranging from 36.4 % to 78.7%. The highest predictive accuracy was obtained for electrical conductivity (EC) (R^2 = 78.7%) followed by silt content (R^2 = 67.1%), and soil pH (R^2 = 57.4%). Moderate performance was observed for P, SOC, exchangeable K, and clay content. The generated spatial soil property maps indicate moderately favourable soil conditions for tea cultivation, characterized by acidic environments (pH 4.2–4.8) and non-saline conditions. These maps provide a valuable decision-support tool for land-use planning and sustainable tea management.

Key words: Digital soil mapping, Random Forest, SCORPAN, Soil properties, Tea cultivation, GIS

Corresponding Author:
Name: Finias F. Mwesige
Email: finiasfidelis@gmail.com

1. Introduction

Soil serves as the fundamental basis for terrestrial ecosystems, providing essential support for global agriculture, biodiversity, and ecosystem services [1, 2]. In the modern era of sustainable land management, there is an escalating global demand for high-

resolution, quantitative, and spatial soil information to guide evidence-based decision-making [1, 3, 4]. While significant efforts have been made to compile global soil databases, spatial soil information remains fragmented and limited in many regions, particularly across the African continent [4, 5, 6]. In sub-Saharan Africa, approximately 80% of arable land suffers from physical soil problems or low fertility, yet the available information to address these challenges is often outdated or restricted to specific zones of interest [5]. Consequently, bridging this gap should become a critical priority for ensuring food security and achieving sustainable agricultural intensification.

Traditionally, soil information is generated through conventional soil mapping (CSM), which relies on manual field surveys and qualitative expert knowledge to delineate discrete soil units [1, 7]. However, CSM is frequently criticized for being labour-intensive, expensive, time-consuming, and difficult to reproduce, as the conceptual models used by surveyors are rarely formalized [3, 7, 8]. In contrast, digital soil mapping (DSM) has revolutionized soil science by combining traditional techniques with modern computing, geographic information systems (GIS), and remote sensing. DSM offers distinct advantages over conventional methods, including increased efficiency, objective reproducibility, and the ability to provide quantified measures of prediction uncertainty [5, 7]. Furthermore, digital methods are significantly more cost-effective for large-scale applications and produce higher-resolution outputs that better simulate the continuous variation of soil properties across a landscape [5, 8].

The digital revolution in DSM and soil science at large is largely driven by advances in machine learning (ML), which have enabled the modelling of complex, non-linear relationships between soil properties and environmental covariates [1, 6, 9]. Globally,

DSM has transitioned from academic research to operational use, with algorithms such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), Random Forest (RF) and other tree-based models being widely tested [3, 6, 10]. While SVM can model non-linear relationships, its performance is often sensitive to hyperparameter selection and prone to overfitting [6, 9]. Similarly, ANNs are robust for complex datasets but typically require very large sample sizes to achieve stability [1, 6]. Random Forest (RF) has emerged as a premier algorithm in the DSM community due to its robustness, versatility, and superior predictive performance in both regression and classification tasks [1, 5, 10, 11]. RF is particularly valued for its ability to handle high-dimensional data, its resistance to noise, and its capacity to avoid overfitting through an ensemble of independent decision trees [11, 12, 13]. Generally, comparative studies consistently demonstrate that RF frequently outperforms other ML algorithms and linear regression models, justifying its status as a benchmark tool for spatial soil prediction.

In Tanzania, the availability of detailed spatial soil information remains inadequate and fragmented [12, 13, 14]. The current national soil data largely consists of legacy maps produced at coarse scales, which lack the resolution necessary for practical farm-level planning or site-specific management. While national initiatives such as the National Forest Resources Monitoring and Assessment (NAFORMA) and the Soil Information System for Tanzania (SISTAN) have made strides in documenting soil carbon and establishing digital records, these efforts are often limited by budgetary constraints or focused on broad ecological assessments [14, 15, 16].

Early initiatives toward a computerized soil management system in Tanzania were marked by soil scientists and researchers in the country transitioning from manual

surveys to computerized approaches through the establishment of the Soil Information System for Tanzania (SISTAN) [16, 17], later Tanzania's Soil Information System (SIS). Currently, SIS is primarily driven by the Tanzania Soil Information Service (TanSIS), established around 2018 which integrates modern DSM. Early initiatives were also demonstrated by Kimaro et al. [15] through utilization of automated land evaluation system (ALES) in soil mapping and land evaluation. Current initiatives include DSM using machine learning techniques (especially Random Forest) to update legacy maps [4, 13, 18] and evaluation of land suitability for rice [12], and sisal [10]. These initiatives also include national soil organic carbon mapping [14], and the use of remote sensing and GIS for site-specific agricultural intensification [19]. However, many regions still rely on conventional survey data, which does not capture many details of soil variability over the landscape.

This data gap is particularly acute for major cash crops like tea, which require precise soil conditions, such as a pH between 4.5 and 5.5 and well-drained fertile soil, to optimize productivity [20]. Literature indicates that in Tanzania, tea research and soil mapping have historically given limited attention to the systematic mapping of soils within existing tea plantations and potential tea-growing areas. Meanwhile, smallholder farmers in tea-producing areas such as Tarime District report yields as low as 0.23 t/ha [21], far below the national potential of about 5 t/ha. This shortfall can be attributed to the absence of baseline land-use and soil information needed to support evidence-based decision-making. To address these production constraints, this study applies Random Forest–based DSM in the Ganyange area of Tarime District to generate detailed spatial soil information for use particularly in multi-criteria land-suitability evaluation for tea cultivation. This will optimize agricultural resource allocation and support the

sustainable expansion of tea production in Tanzania, in alignment with local strategic goals and national efforts to boost the agricultural sector.

2. Methodology

2.1. Study area

The study covered Ganyange Ward in Tarime District located in Mara Region, Tanzania. The geographical location is between latitudes 9847283—9855694 S and longitudes 673079—681185 E and covers an area of 2,677.96 ha (**Fig. 1**). The altitude of the study area ranges between 1590 m to 1850 m above sea level. The area receives annual rainfall ranging from 1,177 to 1,415 mm, with a mean annual total of approximately 1,280 mm. Maximum monthly temperatures range from 25 °C to 27 °C, while minimum monthly temperatures range between 15 °C and 16 °C. The landscape comprises of undulating, rolling and plain landforms. According to International Soil Reference and Information Centre (ISRIC) [22], the dominant soil type is Nitisols.

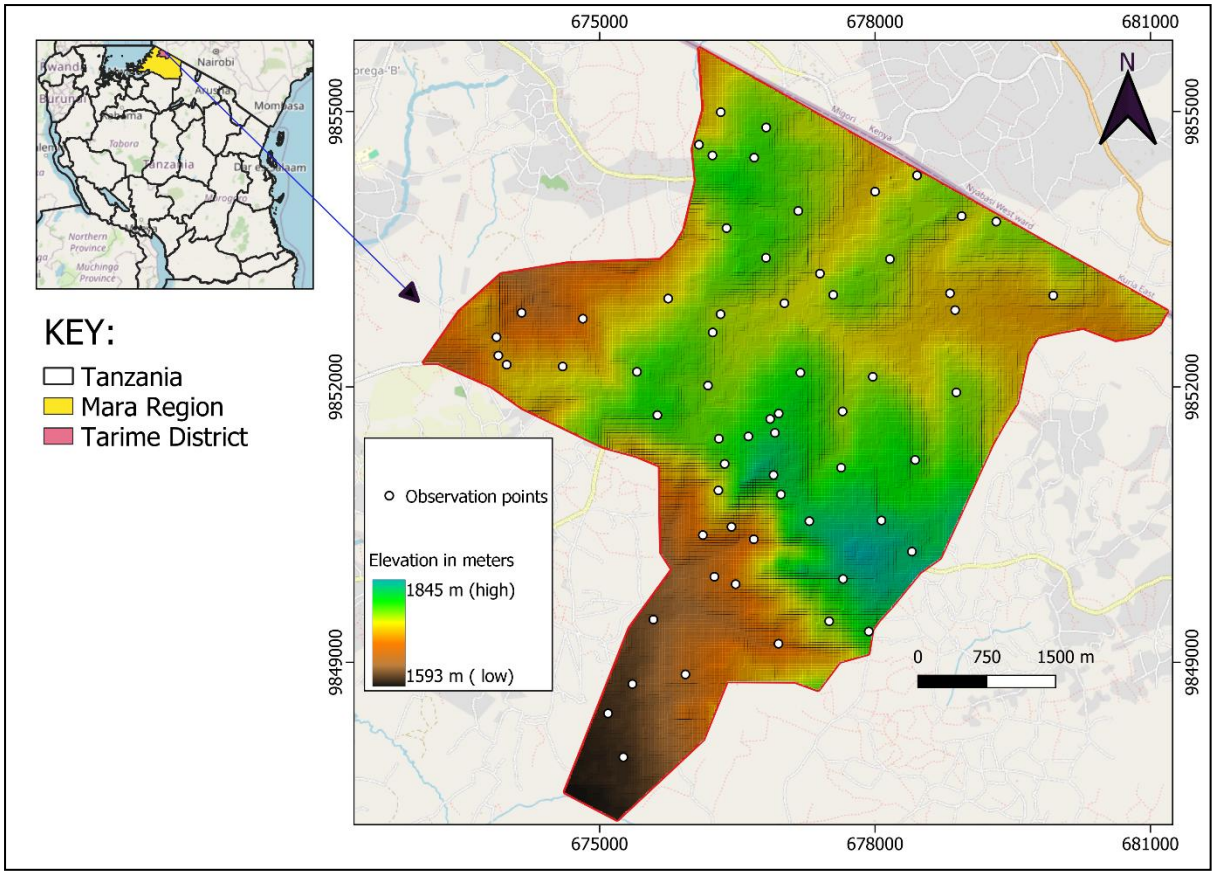


Fig. 1. Spatial distribution of observation sites relative to elevation in the study area.

2.2. Soil survey and sampling procedures

The conditioned Latin Hypercube Sampling (cLHS) approach, following the procedures by Brus [23], was applied to generate representative sampling locations capable of capturing the maximum variability of soils across the study landscape. The sampling design was guided by a set of environmental covariates, including topographic and spectral variables, to ensure that the selected locations adequately represented the range of environmental conditions influencing soil formation.

Using a handheld Global Positioning System (GPS) device, a total of 16 soil profile sites were identified and excavated to a depth of 200 cm, or until a restrictive layer was

encountered, in order to expose the genetic horizons. Horizon-wise morphological characteristics including depth, colour, structure, texture, coarse fragments, consistence, the presence or absence of concretions, nodules and mottles, the occurrence of cutanic coatings, and evidence of biological activity were described in accordance with the Food and Agriculture Organization of the United Nations (FAO) Guidelines for Soil Description [24]. Soil samples from each identified genetic horizon were subsequently collected for laboratory analysis. In addition to the profile observations, supplementary surface sampling locations were determined using a handheld GPS device. At these locations, 48 soil samples were collected using an auger and submitted for laboratory analysis to complement the profile data.

2.3. Laboratory analysis

The collected soil samples were air-dried, passed through a 2-mm sieve, and preserved for subsequent determination of their physical and chemical characteristics. Soil particle-size distribution was determined using the Bouyoucos hydrometer technique [25]. Soil reaction (pH) was measured potentiometrically in a soil–water suspension prepared at a 1:2.5 ratio [26], while electrical conductivity (EC) was measured by conductivity meter in a 1:2.5 soil: water suspension by a method described by Rhoades [32]. Cation exchange capacity (CEC) was evaluated by saturating the soil with neutral 1 M ammonium acetate (NH_4OAc), followed by displacement of the adsorbed NH_4^+ ions using 1 M potassium chloride (KCl). The displaced ammonium was then quantified through the Kjeldahl distillation procedure to estimate soil CEC [27]. Soil organic carbon content was analysed using the Walkley and Black wet oxidation method [28]. Total nitrogen was determined by the Kjeldahl digestion method as described by Bremner [33]. Available phosphorus was extracted using the Bray-I method for acidic soils [29], while the Olsen extraction

method was applied for alkaline soils [30]. Exchangeable base cations were extracted with ammonium acetate, after which calcium (Ca) and magnesium (Mg) concentrations were measured using an atomic absorption spectrophotometer [27]. Exchangeable potassium (K) and sodium (Na) in the extract were quantified by flame photometry [31]. Summary statistics for the soil laboratory results is presented in **Table 1.**

Table 1. Summary statistics for soil laboratory results.

Parameter	**Unit**	**Min**	**Max**	**Mean**	**SD**
Soil pH		4.11	4.9	4.56	0.16
EC	$dS\ m^{-1}$	0.02	0.25	0.11	0.05
SOC	%	1.03	5.76	3.55	0.92
N	%	0.09	0.27	0.17	0.03
P	$mg\ kg^{-1}$	0.76	11.47	4.54	1.92
K	$cmol(+)\ kg^{-1}$	0.15	2.5	0.77	0.41
Ca	$cmol(+)\ kg^{-1}$	0.27	12.28	7.39	1.59
Mg	$cmol(+)\ kg^{-1}$	0.13	4.2	2.53	0.59
CEC	$cmol(+)\ kg^{-1}$	28.56	79.05	53.14	8.27
Sand	%	32	72	50	7.9
Silt	%	14	43	23	5.3
Clay	%	10	42	27	6.0
Na	$cmol(+)\ kg^{-1}$	0.1	2.05	0.28	0.26

pH=soil pH, EC = electrical conductivity, SOC = soil organic carbon, N = total nitrogen, P = available phosphorus, K = exchangeable potassium, Ca =exchangeable calcium, Mg = exchangeable magnesium, CEC = cation exchange capacity, Na = exchangeable sodium

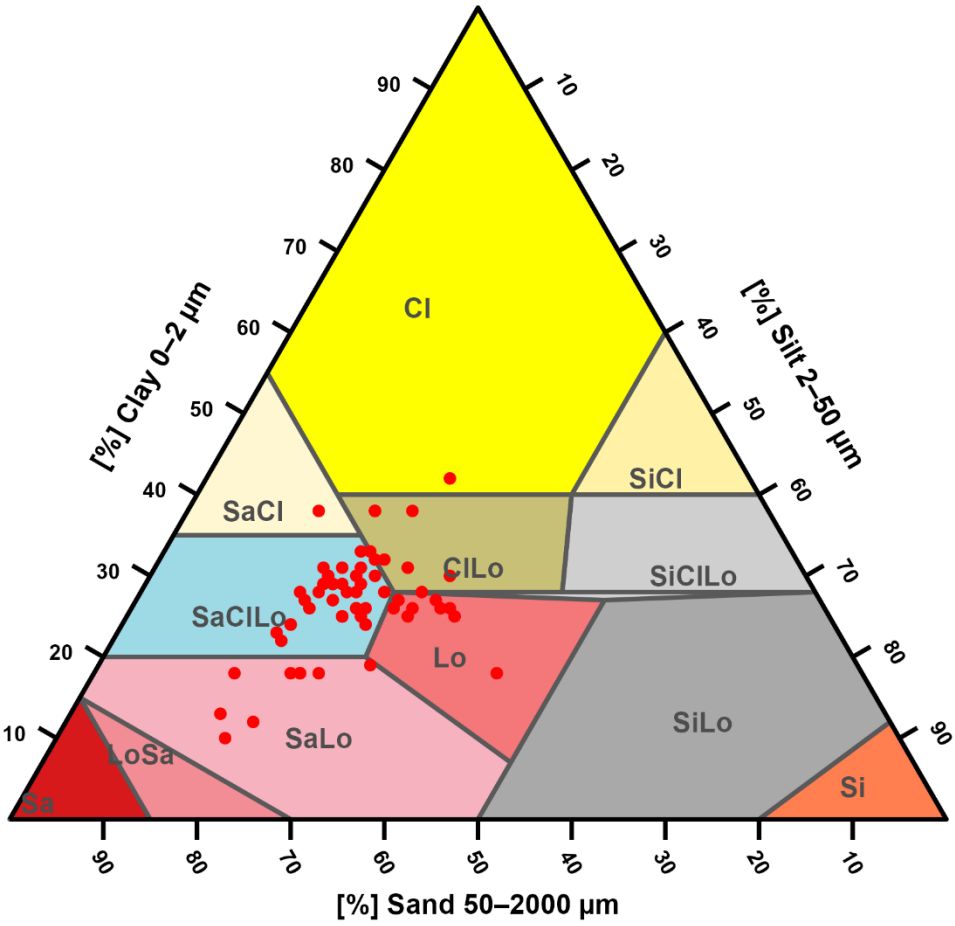


Fig. 2. Spatial distribution of soil texture classes for 0–30 cm splined data, illustrated on the USDA soil texture triangle. The positions of observations are represented by red dots. Sa (sand), LoSa (loamy sand), SaLo (sandy loam), Lo (loam), SiLo (silt loam), Si (silt), SaClLo (sandy clay loam), ClLo (clay loam), SiClLo (silty clay loam), SaCl (sandy clay), SiCl (silty clay), and Cl (clay).

2.4. Predictive mapping of soil properties

2.4.1. Prediction model

Random Forest (RF) is an ensemble machine-learning algorithm that constructs multiple decision trees to model complex, non-linear relationships between target soil variables and environmental covariates. Prior to modelling, soil property data were log-transformed to reduce skewness; following RF prediction, outputs were back-transformed to the original scale for spatial mapping and interpretation. The generally observed weak relationships between soil properties and environmental covariates reflect intensive agricultural disturbance, which has disrupted natural soil–landscape relationships. To improve prediction reliability, selected soil properties were therefore included as covariates for predicting other soil properties where significant correlations and strong pedological relationships existed. RF models were trained using SCORPAN-based covariates to predict the spatial variability of key soil properties and to identify influential predictors, including terrain-derived variables, spatial indices, spatial soil information, and vegetation indices.

RF modelling was implemented in R using the ranger package, while spatial cross-validation was conducted using the caret Applications for Spatial–Temporal Models (CAST) framework to ensure robust model evaluation. Soil sample data were converted into spatial objects using geographic coordinates (x, y) and an appropriate projected coordinate reference system. Spatial dependence was addressed by generating four spatially disjoint folds using a block-based cross-validation approach, whereby entire spatial blocks were withheld during model training, minimizing information leakage caused by spatial autocorrelation. This procedure resulted in approximately 75% of the

samples being used for training and 25% for validation in each spatial cross-validation fold. Model performance was evaluated across these folds, followed by interpretation of permutation-based variable importance to identify key predictors of the target soil properties. After validation, a final RF model was trained using all available samples and optimized parameters to generate spatially explicit soil property predictions.

2.4.2. Evaluation of model performance

Model performance was evaluated using spatial cross-validation to ensure unbiased assessment under spatial dependence. Prediction accuracy was quantified using three complementary metrics: root mean square error (RMSE) determined using **Eq. 1**, mean absolute error (MAE) calculated using **Eq. 2**, and the coefficient of determination (R^2) computed by **Eq.3**. RMSE was used as the primary indicator of overall error magnitude, while MAE provided a robust measure less sensitive to outliers. R^2 quantified the proportion of variance in observed soil properties explained by the model. For each spatial fold, RF models were trained on spatially independent training subsets (≈75% of samples) and evaluated on withheld validation samples (≈25%), and performance metrics were computed accordingly. Final model performance was summarized using the mean and standard deviation of RMSE, MAE, and R^2 across all spatial cross-validation folds, providing a reliable estimate of predictive performance.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2} \qquad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|yi - \hat{y}i| \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{\sum_{i=1}^{n}(yi - \bar{y})^2} \quad (3)$$

where yi is the observed value, $\hat{y}i$ is the predicted value, $\bar{y}$ is the mean of observed values, and n is the number of observations.

2.5. Pre-processing of environmental covariates for prediction of soil properties

A set of environmental covariates comprising of remote sensing–derived reflectance data, Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) derived topographic variables, and kriged soil raster maps was employed in the current study **(Table 2).** Multispectral imagery from Sentinel-2A, obtained from the European Space Agency Copernicus Data Space [34], provided the primary spectral covariates, including individual bands in the blue (Band 2), green (Band 3), red (Band 4), near-infrared (NIR; Band 8), cirrus (Band 10), shortwave infrared 1 (SWIR-1; Band 11), and shortwave infrared 2 (SWIR-2; Band 12) regions. In addition to the raw spectral bands, a set of spectral indices was derived to enhance soil and vegetation characterization. These indices comprised the bare soil index (BSI), clay mineral ratio (CMR), coloration index (CI), ferrous iron index (FII), redness index (RI), coloration index (CI), grain size index (GSI), calcium-related normalized difference index (CalCI), normalized difference vegetation index (NDVI), soil adjusted vegetation index (SAVI), land use land cover

(LULC), optimized soil adjusted vegetation index (OSAVI), and soil enhancement ratio (SER).

Table 2. Environmental covariates used as predictors for digital soil mapping in this study.

Source	Covariate data	Formula/Definition	Resolution
Sentinel 2A	$B2$, $B3$, $B4$, $B8$, $B10$, $B11$ & $B12$ spectral bands	$B2$ = Blue, $B3$ = Green, $B4$ = Red, $B8$= NIR, $B10$= Cirrus, $B11$ = Shortwave Infrared 1—SWIR-1, and B12 = Shortwave Infrared 2—SWIR-2	10 m
	Bare Soil Index (BSI)	$BSI = \frac{(B11 + B4) - (B8 + B2)}{(B11 + B4) + (B8 + B2)}$	10 m
	Clay Mineral Ratio (CMR)	$CMR\ \frac{B11}{B12}$	10 m
	Coloration Index (CI)	CI = $\frac{B4}{B2}$	10 m
	Ferrous Iron Index (FII)	$FII = \frac{B11}{B8}$	10 m

Grain Size Index (GSI)	$GSI\ \frac{B4 - B2}{B4 + B2 + B3}$	10 m
Normalized Difference Calcareous Index (CalCI)	$CalCI = \frac{B11 - B12}{B11 + B2}$	10 m
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{B8 - B4}{B8 + B4}$	10 m
Redness Index (RI)	$RI = \frac{B4^2}{B2\ X\ B3}$	10 m
Soil Adjusted Vegetation Index (SAVI)	$SAVI = \frac{(B8 - B4)}{(B8 + B4 + L)}\ X\ (1 + L)$ where L is the soil brightness correction factor (commonly $L = 0.5$)	10 m

	Land Use Land Cover (LULC)		10 m
	Optimized Soil-Adjusted Vegetation Index (OSAVI)	$OSAVI = \frac{B8 - B4}{B8 + B4 + 0.16}$	10 m
	Soil Enhancement Ratio (SER)	$SER = \frac{B11}{B8}$	10 m
DEM(SRTM)	Elevation		30 m
	Slope		30 m
	Aspect		30 m
	Slope Length		30 m
	Topographic Wetness Index (TWI)		30 m
	Plan Curvature		30 m
	Total Curvature		30 m
Kriged soil maps	Soil pH		10 m

Soil Organic Carbon (SOC)	10 m
Potassium (K)	10 m
Calcium (Ca)	10 m
Magnesium (Mg)	10 m
Sodium (Na)	10 m
Electrical conductivity (EC)	10 m
Cation exchange Capacity (CEC)	10 m
Silt content	10 m
Sand content	10 m

Topographic predictors were extracted from the U.S. Geological Survey (USGS) 30-m SRTM DEM [35], using the System for Automated Geoscientific Analyses–Geographic Information System (SAGA-GIS) processing framework. The derived topographic data included elevation, slope, aspect (As), slope-length (LS) factor, topographic wetness index (TWI), drainage, plane curvature and total curvature. In addition, Ordinary Kriging (OK)–derived ancillary soil raster layers of correlated soil properties showing strong pedological relationships with the target soil property were incorporated alongside topographic and spectral covariates as additional predictors in the modelling process. All environmental covariates were standardized to a common coordinate reference system, spatial extent, and a uniform spatial resolution of 10 m using ArcGIS 10.4 software. **Table 3** presents a summary of processed soil, topographic, and vegetation covariates used as predictors of key soil properties for tea growth in the study area.

Table 3. Summary statistics for soil, topography, and spectral covariates.

Factor	Unit	Min	Max	Mean	SD
Soil					
pH		4.23	4.72	4.53	0.10
EC	$dS\ m^{-1}$	0.06	0.22	0.12	0.03
SOC	%	2.19	4.89	3.57	0.54
N	%	0.13	0.21	0.17	0.01
P	$mg\ kg^{-1}$	2.80	8.74	4.27	0.89
K	$cmol(+)\ kg^{-1}$	0.42	1.51	0.78	0.19
Ca	$cmol(+)\ kg^{-1}$	4.22	9.31	7.44	0.92
Mg	$cmol(+)\ kg^{-1}$	1.15	3.36	2.51	0.23
CEC	$cmol(+)\ kg^{-1}$	37.67	60.91	52.43	3.47
sand	%	12.18	62.61	46.14	6.47
silt	%	16.49	36.33	23.82	4.38
Na	$cmol(+)\ kg^{-1}$	0.11	0.70	0.27	0.11

Topography					
DEM	*m*	1593	1845	1742	50
slope	*%*	0.00	48.63	8.50	5.55
aspect	*degree*	0.02	360.00	199.46	106.47
Spectral data					
NDVI		0.015	0.897	0.606	0.105
SAVI		0.004	0.655	0.367	0.073
OSAVI		0.006	0.663	0.411	0.076
SER		0.011	0.421	0.217	0.045
BSI		-0.442	0.303	-0.026	0.097
CalCL		0.118	0.757	0.550	0.041
CI		-1.069	0.727	0.401	0.094
CMR		1.022	2.456	1.563	0.153
FII		0.926	3.023	1.617	0.118
GSI		-0.248	0.420	0.156	0.049
RI		-0.403	0.413	-0.017	0.087

2.6. Model training and validation dataset

Soil data for model training and validation were obtained from field surveys and laboratory analyses. A total of 64 soil samples were collected, comprising 16 surface-horizon samples from representative soil profile pits—harmonized to a common 0–30 cm depth by mass-preserving spline functions [36], and 48 surface soil samples (0–30 cm) collected using an auger. From these samples, laboratory analyses were conducted for soil properties including soil reaction (pH), electrical conductivity (EC), soil organic carbon (SOC), total nitrogen (N), available phosphorus (P), exchangeable base cations (K, Ca, Mg, and Na), cation exchange capacity (CEC), and sand, silt and clay soil textural classes (**Fig. 2**). Descriptive statistics for these soil properties are presented in **Table 1**. During model prediction, these laboratory results were used for model training (75%) and validation (25%), whereas remote sensing spectral indices, topographic attributes, and ancillary spatial soil information served as predictor variables.

3. Results and discussion

3.1. Prediction of soil organic carbon (SOC) using Random Forest

SOC prediction (**Table 4**) exhibited moderate accuracy (R^2 = 44.7%, RMSE = 0.155%, MAE = 0.101%). This level of performance is typical for SOC mapping in cultivated tropical soils, where management practices introduce high small-scale variability [37, 38, 39]. Literature establishes that while RF is robust, its performance in predicting SOC is heavily influenced by sampling density, the quality of auxiliary environmental data, and intrinsic soil variability. Various studies which predicted SOC by RF from environmental predictors, show an R^2 range of approximately 0.06—0.97[37, 38, 39, 40, 41, 42, 43, 44, 45, 46].

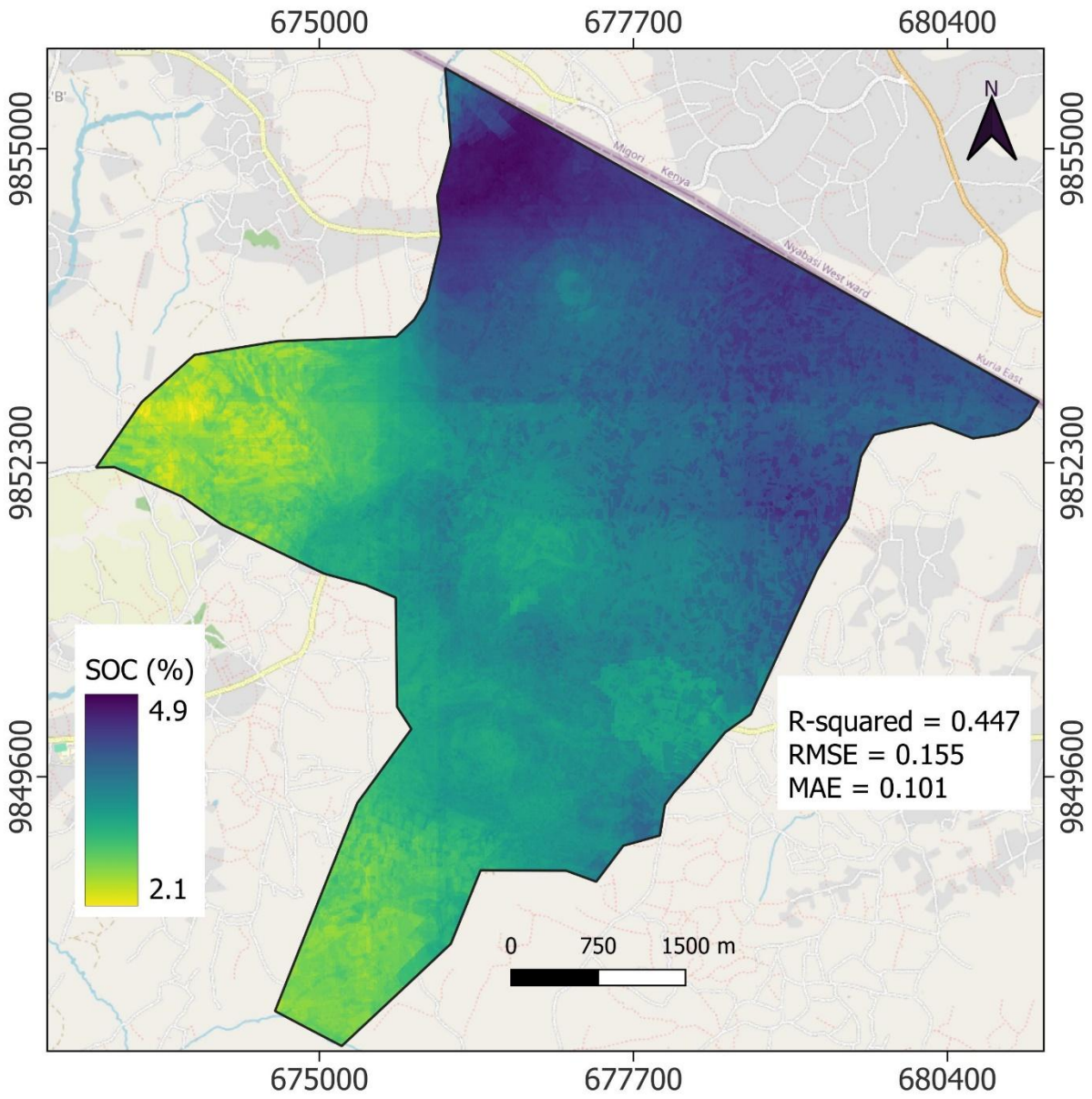


Fig. 3. Prediction map of SOC produced by Random Forest.

Key predictors included CEC, K, pH, DEM, slope, OSAVI, LULC, BSI and clay mineral indices such as CMR (**Fig. 4**). Similar environmental variables and their derivatives have been identified by several studies as key predictors of SOC [38, 41, 44, 45, 46, 47, 48]. The importance of CEC and exchangeable bases is pedologically consistent, as stable organo-mineral complexes enhance carbon stabilization. Correlation analysis confirmed positive

associations between SOC, CEC, and base cations, though the relationships are moderate which could be attributed to fertilizer-driven dissociation of nutrient and carbon pools.

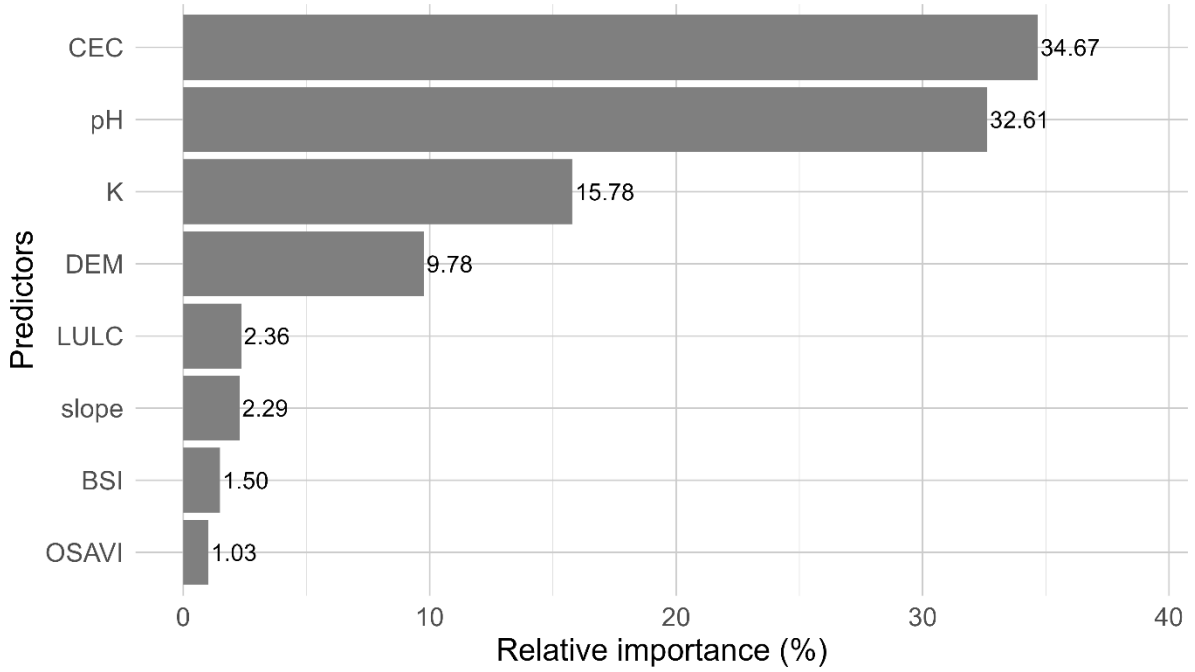


Fig. 4. The relative importance of environmental variables for predicting SOC content.

Predicted SOC (**Fig. 3**) ranges from low to moderate [49]. For tea cultivation, moderate to high SOC is desirable due to its role in nutrient supply, moisture retention, and soil structure [50]. The results indicate that while some areas are favourable, others may benefit from organic matter management strategies to enhance long-term soil quality.

Table 4. Random Forest performance in the prediction of studied soil properties.

Soil property	Metrics of prediction accuracy				Standard deviations of the metrics		
	Unit	**RMSE**	**MAE**	**R^2**	**RMSE - SD**	**MAE - SD**	**R^2 - SD**
silt	*%*	0.124	0.090	0.671	0.015	0.008	0.058
EC	*$dS\ m^{-1}$*	0.019	0.013	0.787	0.006	0.004	0.077
pH		0.018	0.012	0.574	0.009	0.005	0.268
P	*$mg\ kg^{-1}$*	0.257	0.188	0.409	0.116	0.075	0.197
SOC	*%*	0.155	0.101	0.447	0.044	0.022	0.179
K	*$cmol(+)\ kg^{-1}$*	0.135	0.095	0.490	0.026	0.013	0.099
clay	*%*	0.197	0.132	0.364	0.066	0.033	0.231

pH=soil pH, EC=electrical conductivity, P = available phosphorus, SOC = soil organic carbon, K = exchangeable potassium, RMSE=Root Mean Square Error; MAE=Mean Absolute Error; R^2=Coefficient of Determination; SD=Standard Deviation

3.2. Prediction of soil reaction (pH) using Random Forest

Soil pH was predicted with good accuracy (R^2 = 57.4%, RMSE = 0.018, MAE = 0.018) (**Table 4**), demonstrating that RF effectively captured the spatial variability of soil reaction. Key predictors included sand content, EC, DEM, total curvature, SOC, FII, slope and aspect (**Fig. 5**). The influence of sand suggests the role of texture-mediated pH buffering capacity. The role of EC and SOC highlights the influence of fertilizer inputs and organic amendments, which, through nitrification and organic acid production, contribute to soil acidification under intensive cultivation. Terrain variables further modulate pH through leaching and redistribution of base cations in this high-rainfall tropical landscape. Similar levels of performance have been reported in intensively

cultivated soils, where pH responds to both intrinsic soil properties and management practices, and an R^2 range of approximately 0.27—0.89 is demonstrated [41, 43, 44, 48, 51]. Some of these predictors including DEM [43, 44], and Sentinel-2 imagery indices [43] have also been demonstrated as key predictors in the literature.

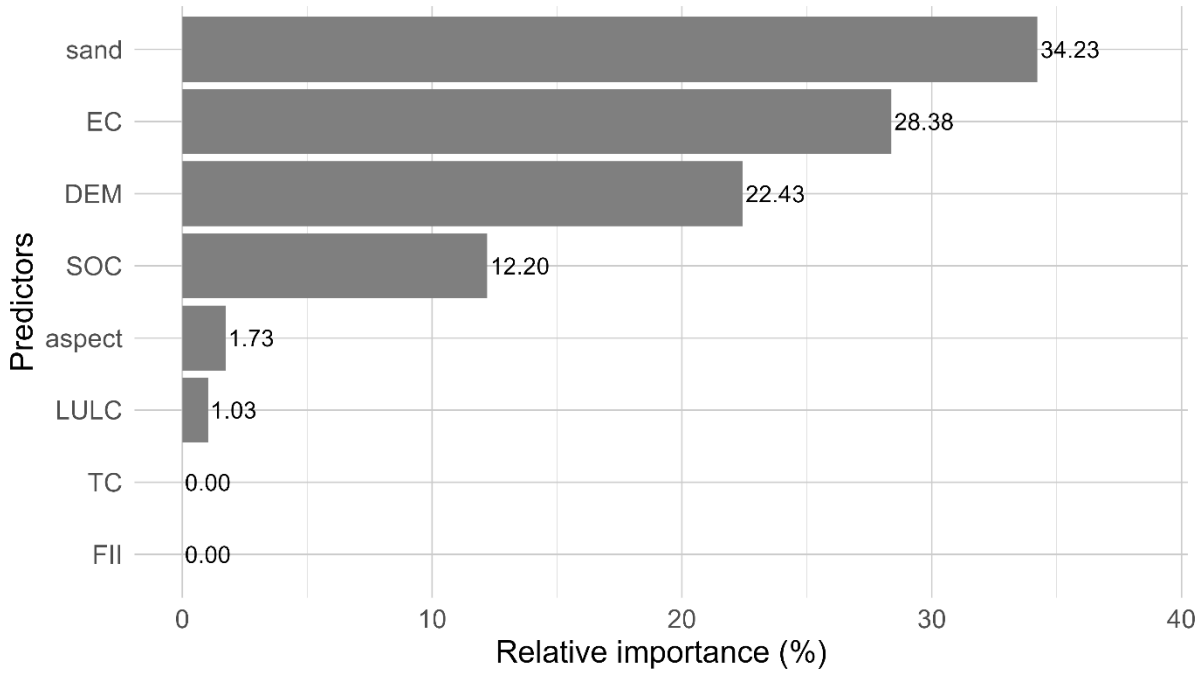


Fig. 5. The relative importance of environmental variables for predicting soil pH.

Predicted pH values (**Table 5**) largely indicate acidic to moderately acidic environment [49]. Literature indicates that pH ranges around 4.5–5.5 are considered optimal, while higher pH values reduce micronutrient availability and tea quality [50, 52, 53, 54, 55, 56, 57, 58]. The RF map (**Fig. 6**) therefore suggest that substantial portions of the study area are well suited for tea cultivation in terms of soil reaction, although localized liming or acidification management may be required where pH deviates from the optimal range.

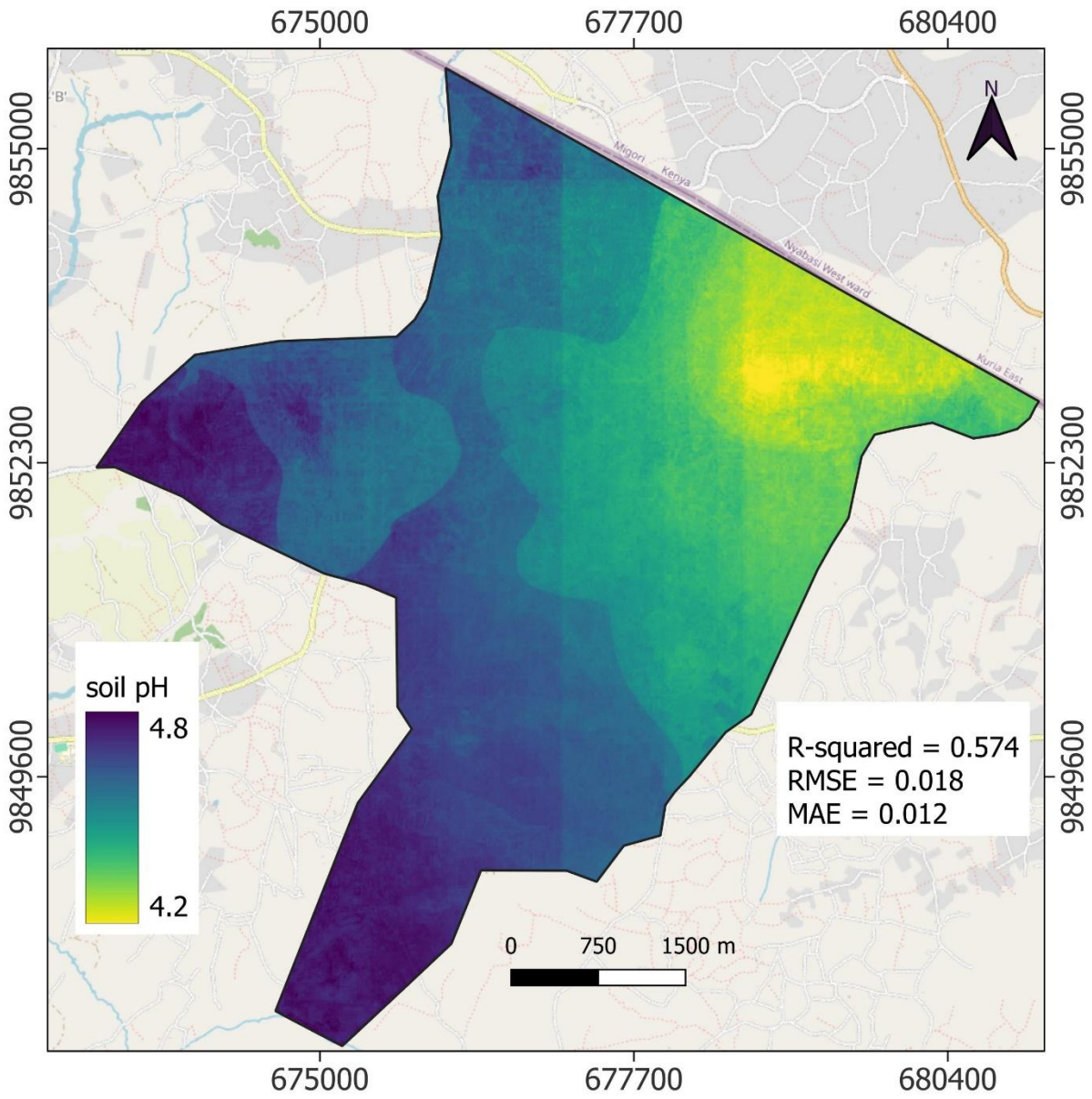


Fig. 6. Prediction map of soil pH produced by Random Forest.

3.3. Prediction of exchangeable Potassium (K) using Random Forest

Exchangeable K was predicted with moderate accuracy (R^2 = 49.0%, RMSE = 0.135 cmol(+) kg^{-1}, MAE = 0.095 cmol (+) kg^{-1}) (**Table 4**). This performance is comparable to other studies in intensively cultivated soils, where K availability is influenced by both soil texture and fertilizer inputs [47, 48, 59].

Important predictors (**Fig. 7**) included silt, Mg, Na, CEC, DEM, and redox-related spectral indices, with much alignment to findings established by Zhang et al. [48], Amare et al. [47], Tamiru et al. [44] and Kaya et al. [60]. Pedologically, the close association between K and CEC reflects the role of exchange sites in retaining monovalent cations [47]. Generally, correlation analyses by Amare et al. [47] confirm positive relationships between K, SOC, and base cations, consistent with nutrient co-accumulation in finer-textured soils.

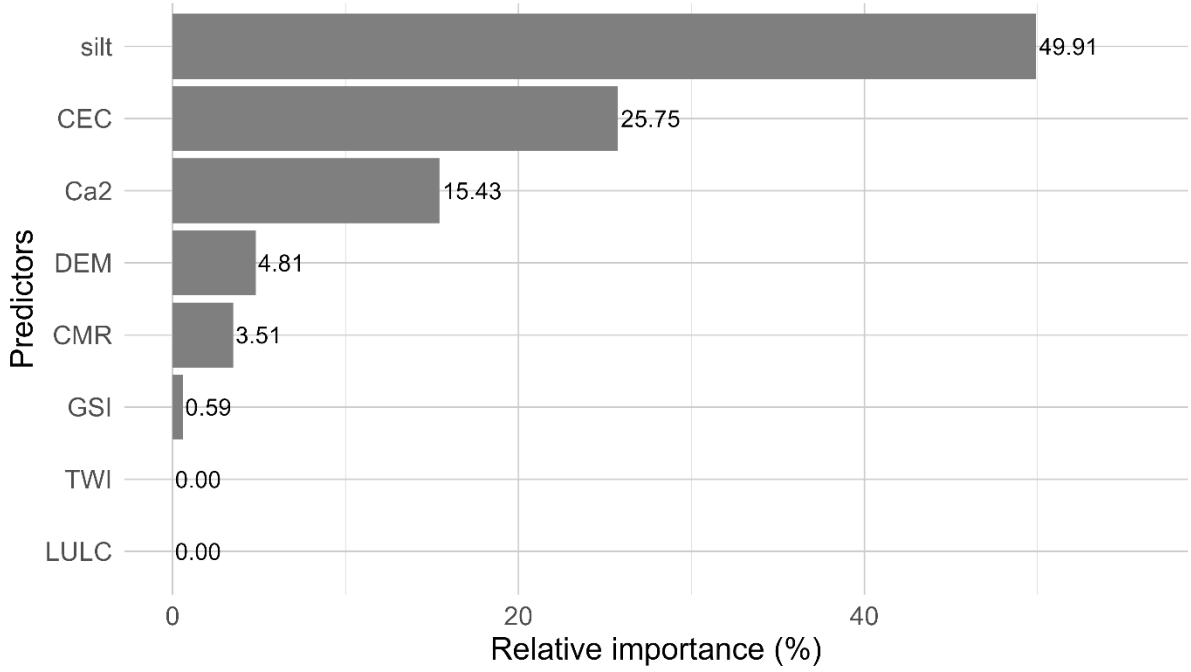


Fig. 7. The relative importance of environmental variables for predicting soil K.

Predicted K levels (**Table 5**) fall within low to medium categories [49]. Adequate K is critical for tea yield and quality, influencing shoot growth and stress tolerance. The RF map (**Fig. 8**) suggests spatial variability in K availability, supporting the need for site-specific nutrient management for tea cultivation in the study area.

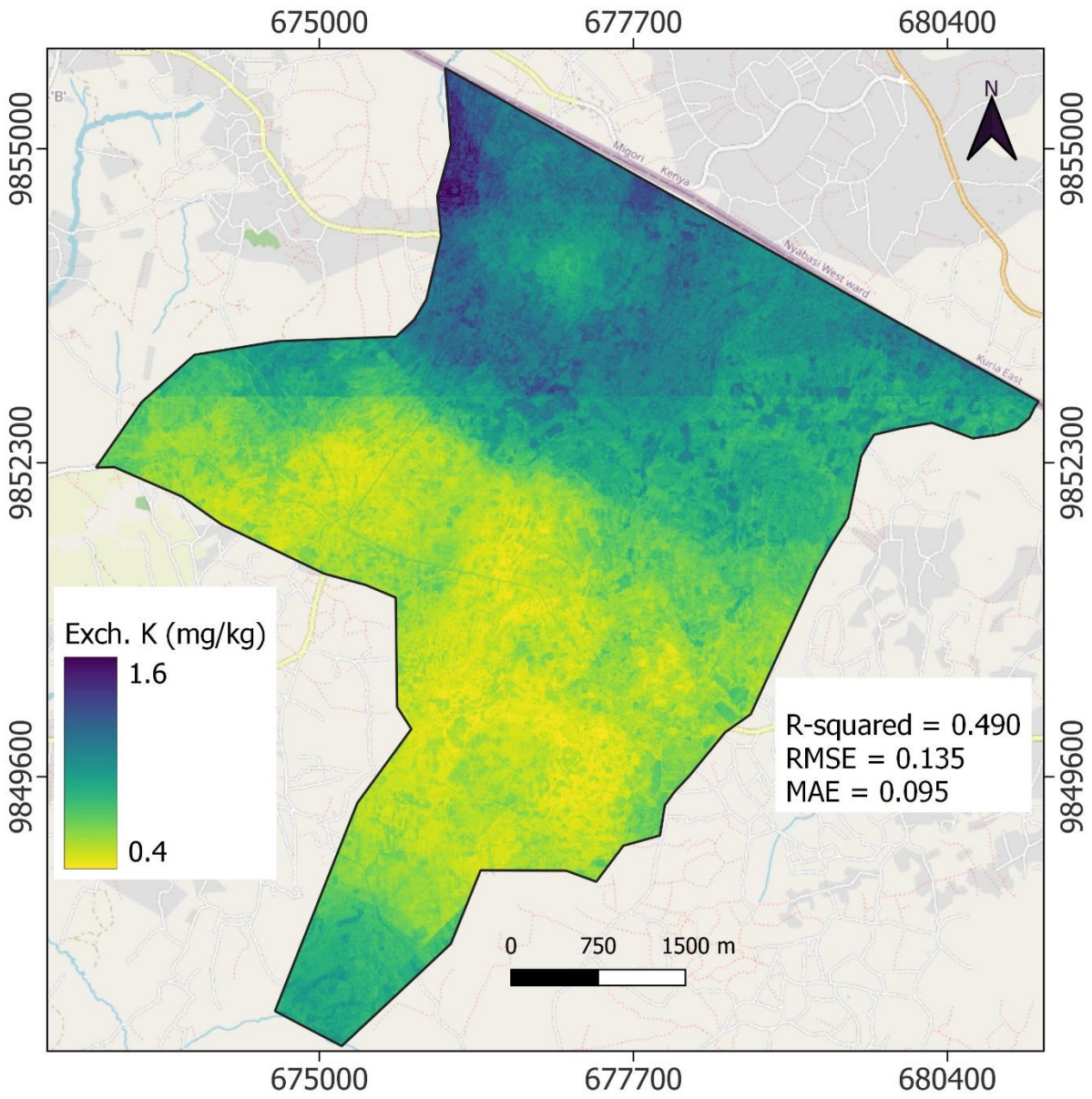


Fig. 8. Prediction map of K produced by Random Forest.

3.4. Prediction of available Phosphorus (P) using Random Forest

Available phosphorus showed moderate predictive performance (R^2 = 40.9%, RMSE = 0.257 $mgkg^{-1}$, MAE = 0.188 $mgkg^{-1}$) (**Table 4**), reflecting greater uncertainty possibly due to limited influence by textural properties. This aligns with the broader literature [47, 51], which frequently reports lower predictability for P due to its strong dependence on localized management, fixation processes, and mineralogical controls.

Important predictors (**Fig. 9**) included silt, Mg, K, CEC, DEM and iron-related indices such as ferrous iron index (FII) and redness index (RI), demonstrating much alignment with findings by Tamiru et al. [44], Makungwe et al. [51], Isong et al. [59] and Amare et al. [47]. Pedologically, the association with fine fractions and CEC is expected, as higher surface area enhances P sorption. However, the correlation analysis revealed weak or negative relationships between P and several soil properties, including N and silt, indicating that anthropogenic fertilizer application and fixation by iron and aluminium oxides could be the ones dominating P dynamics in the study area.

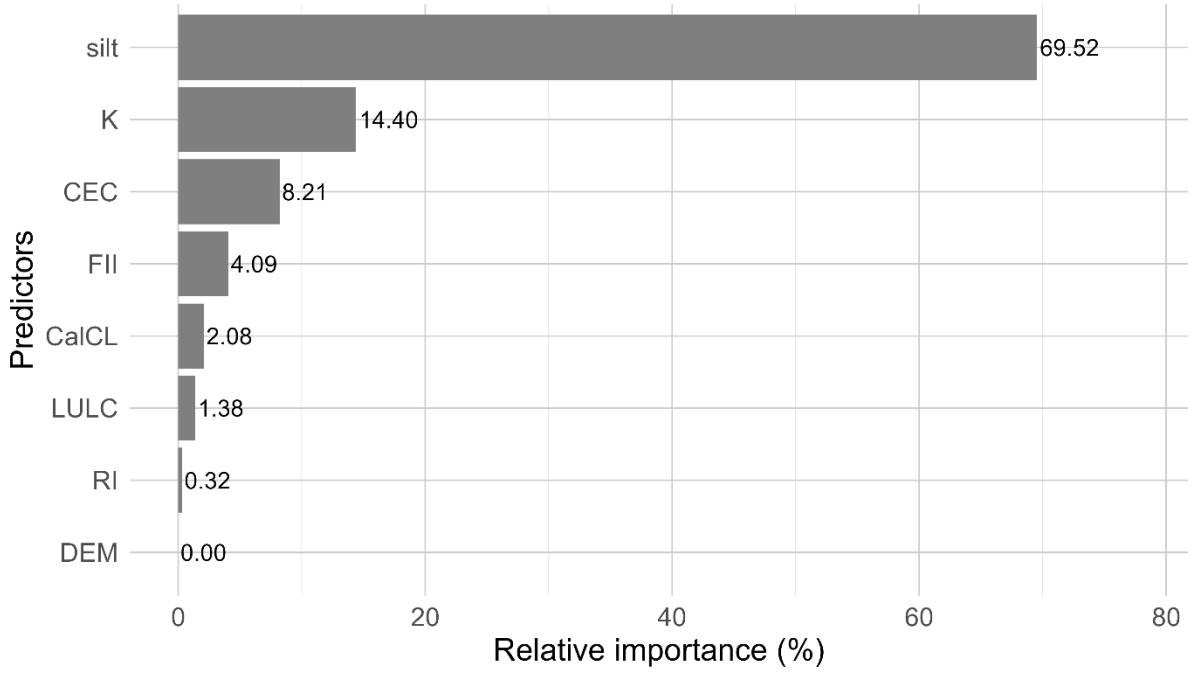


Fig. 9. The relative importance of environmental variables for predicting soil P.

Predicted P levels range from low to moderate (**Fig. 10**), suggesting that P deficiency may constrain tea productivity in parts of the landscape [49]. Tea requires adequate but not

excessive P, and low availability can limit root growth and early establishment [50]. The moderate RF performance implies that while spatial patterns of P can be identified, site-specific soil testing remains essential for precise fertilizer recommendations for tea cultivation.

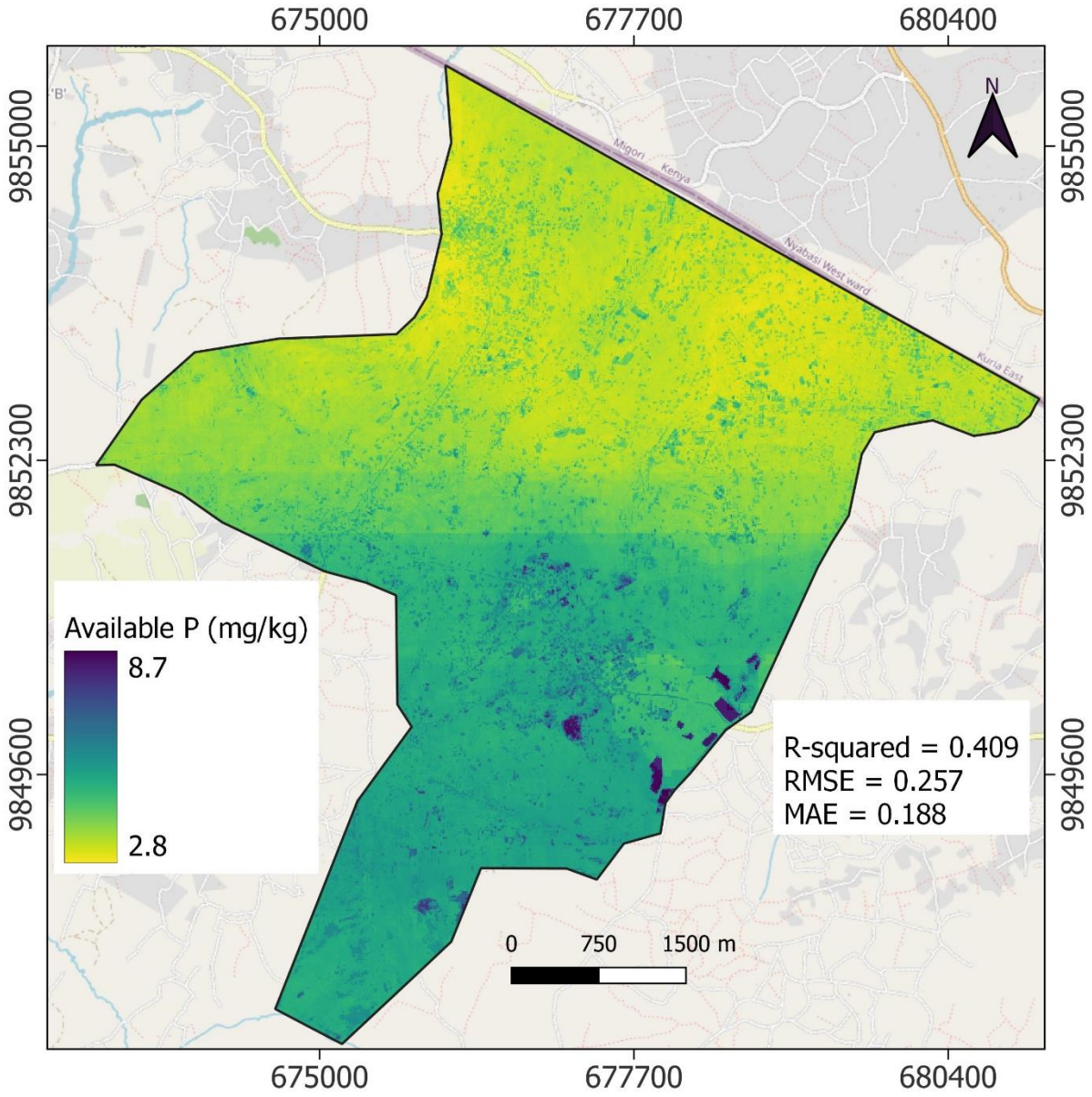


Fig. 10. Prediction map of P produced by Random Forest.

3.5. Prediction of clay percentage using Random Forest

RF showed moderate predictive power for clay content (R^2 = 36.4%, RMSE = 0.197%, MAE = 0.132%) (**Table 4**). Clay is often more difficult to predict than silt due to its complex mineralogical variability and weaker surface spectral response [42, 61]. Key predictors (**Fig. 11**) included Mg, DEM, TWI, SL, aspect, slope, curvature measures such as TC, and mineralogical indices such as clay mineral ratio (CMR) and redness index (RI), showing a greater similarity to findings by Tamiru et al. [44], Naimi et al. [42], and Isong et al. [59]. A study by Mgohele et al. [61] established that the NIR and SWIR2 Sentinel 2 bands are the most critical individual bands for clay prediction, since NIR band is sensitive to moisture and organic matter content closely linked to clay minerals, while the SWIR2 band captures specific absorption characteristics of kaolinite, montmorillonite, and illite, reflecting the complex mineralogical variability inherent in clay.

Zhang et al. [48] asserted that exchangeable Mg and other cations are heavily influenced by higher organic matter and clay content which promotes the adsorption of these base cations by soil colloids. In this regard, the importance of Mg and CMR in the present study could be reflecting the linkage between clay minerals and exchangeable Mg in highly weathered tropical soils typical of the study area. On the other hand, compared to silt, correlation analysis showed weaker associations between clay and other chemical properties, indicating that clay exerts a slighter control on nutrient dynamics in the studied soils.

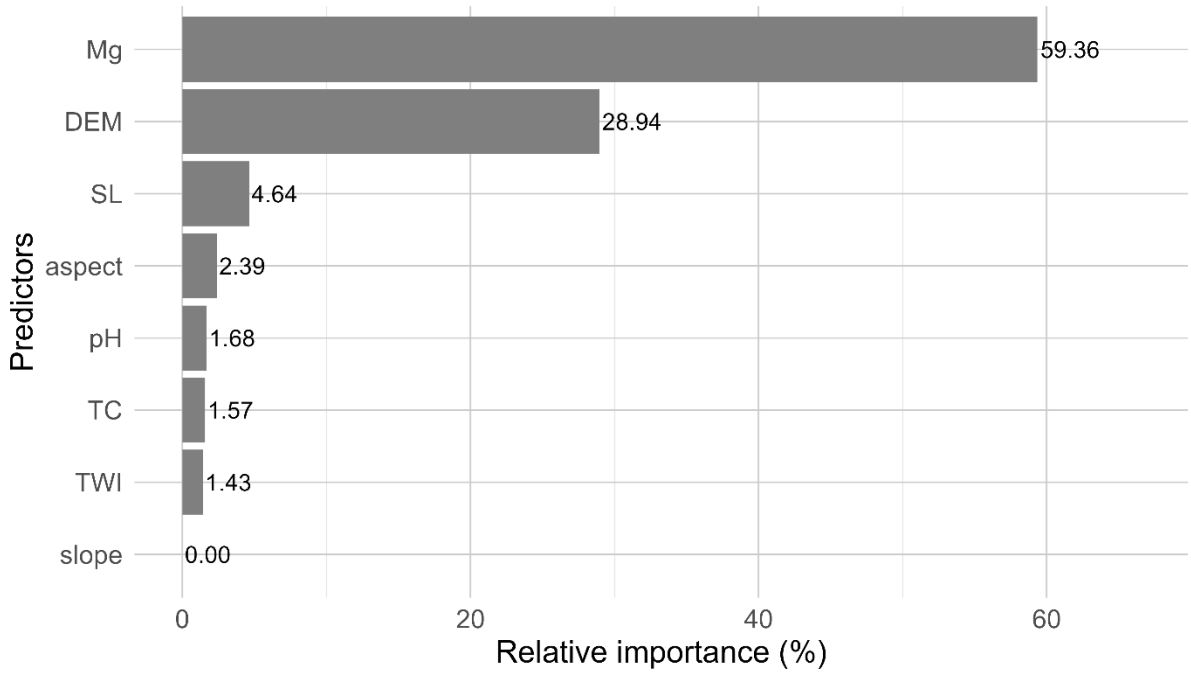


Fig. 11. The relative importance of environmental variables for predicting clay content

Predicted clay contents (**Table 5**) correspond mainly to loamy to clay-loam textures, which are generally suitable to moderately suitable for tea cultivation [52, 54, 55, 56, 57, 58]. Excessively high clay contents can restrict drainage, but moderate clay enhances nutrient retention and moisture supply. The RF-predicted clay map (**Fig. 12**) therefore provides useful, though moderately uncertain, input for tea land suitability evaluation.

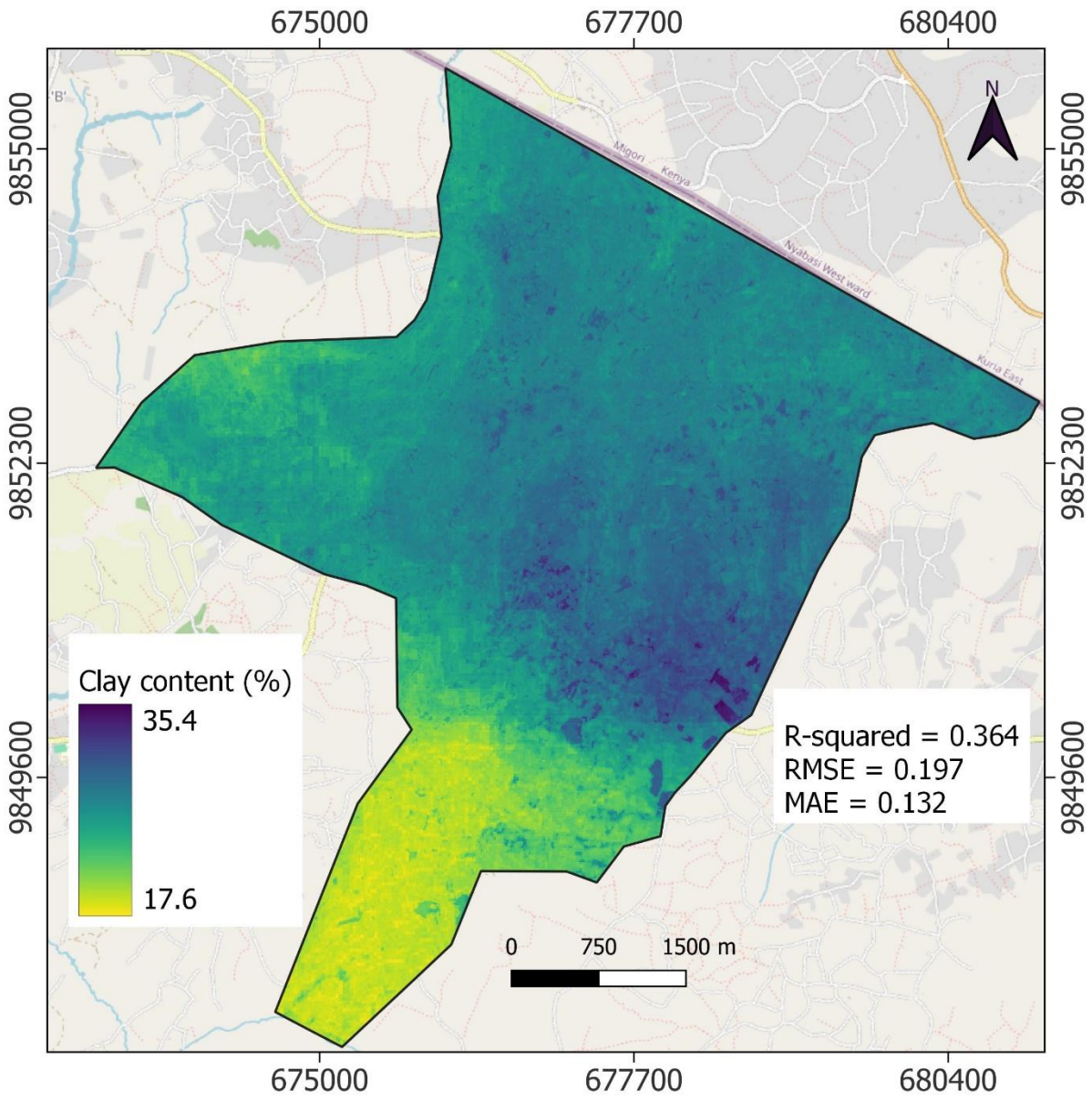


Fig. 12. Clay percentage map predicted by Random Forest

3.6. Prediction of silt percentage using Random Forest

The RF model showed better predictive power for silt content (R^2 = 67.1%, RMSE = 0.124%, MAE = 0.090%) (**Table 4**), indicating strong agreement between observed and predicted values and relatively low prediction uncertainty. Comparable studies in cultivated tropical landscapes have reported similarly high predictability for silt due to its

strong linkage with topographic redistribution processes and soil moisture dynamics, which are well captured by terrain derivatives and spectral indices [47].

Variable importance analysis (**Fig. 13**) identified EC, Ca, Na, DEM, TWI, slope, SER, and grain-size–related indices (e.g. grain size index—GSI , bare soil index—BSI) as the most influential predictors, similarly to findings by Kaya et al. [60], Mgohele et al. [61], Naimi et al. [42], Tamiru et al. [44], Amare et al. [47], and [48]. This pattern is pedologically consistent, as silt-rich soils often exhibit higher nutrient retention and ionic strength than sandy soils. The strong correlations between silt and EC, exchangeable bases, SOC, and CEC observed in the correlation analysis further support this interpretation. Amare et al. [47], Naimi et al. [42], and Tamiru et al. [44] asserted that topographic variables are essential for dictating nutrient-retention gradients in high-rainfall tropical landscapes capturing the long-term redistribution of fine particles like silt from high-elevation areas to low-lying plains. This increases predictability of silt content from topographic attributes.

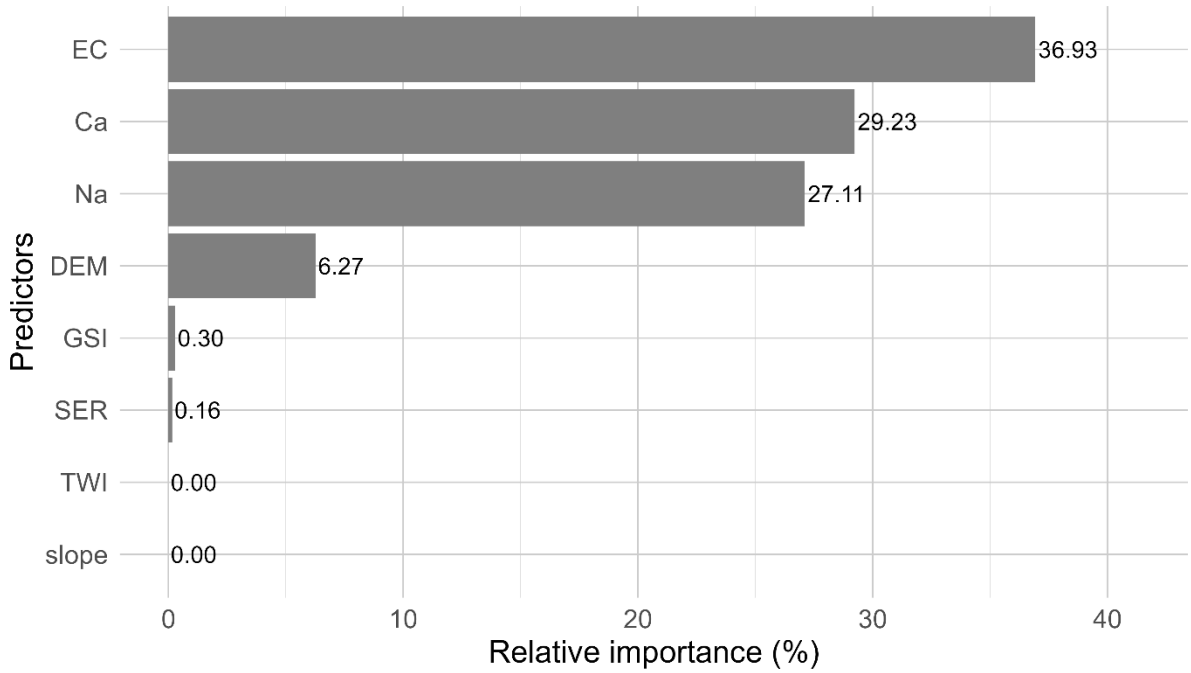


Fig. 13. The relative importance of environmental variables for predicting silt content.

The predicted silt distribution (**Fig. 14**) spans soil textural classes considered suitable to moderately suitable for tea cultivation, such as loam, clay loam, and sandy clay loam [52, 54, 55, 56, 57, 58]. These textures provide favourable water-holding capacity and root aeration for tea. Overall, the strong RF performance for silt confirms that terrain- and texture-regulated soil properties can be robustly mapped using machine-learning approaches in managed tropical soils.

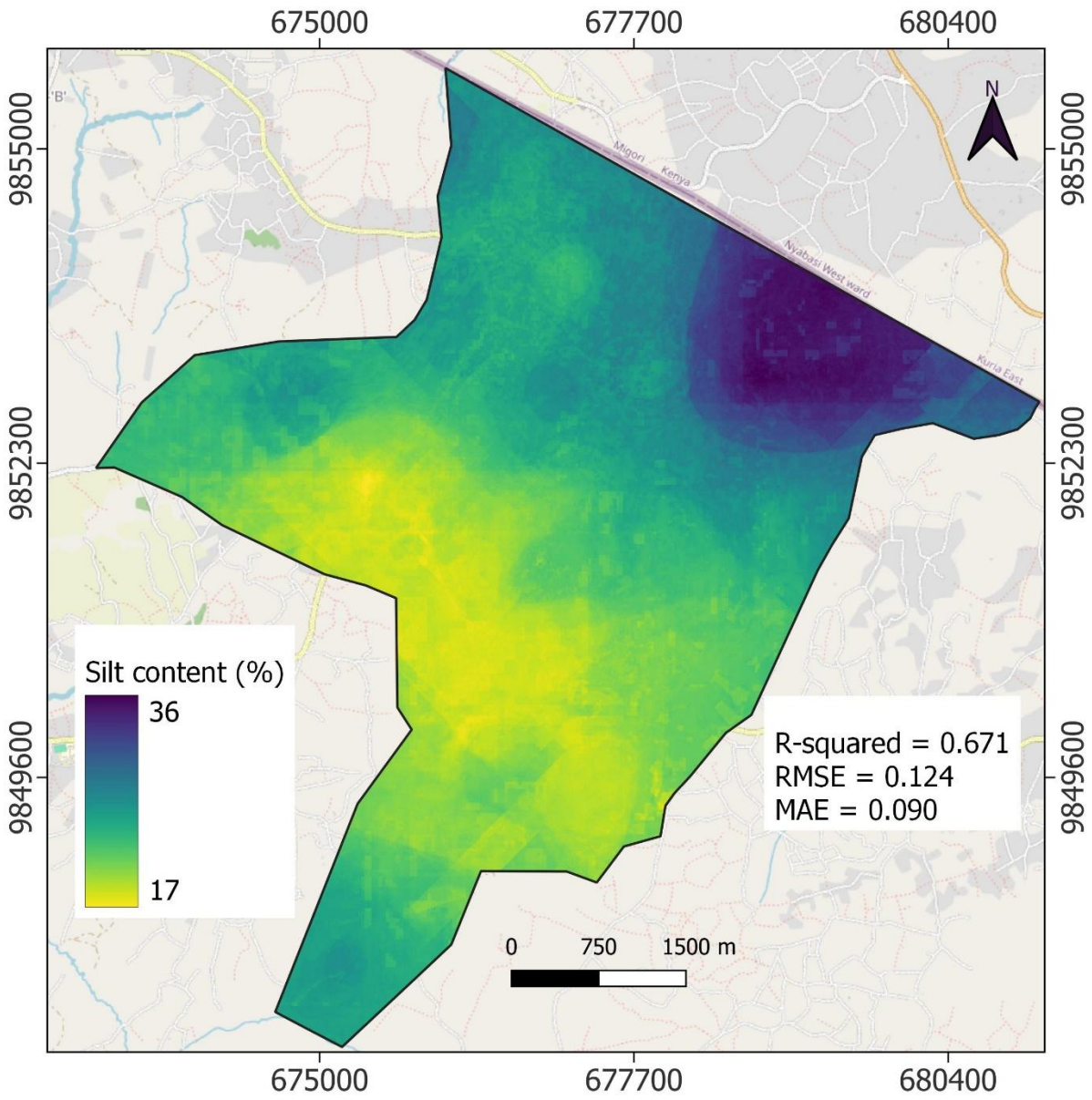


Fig. 14. Silt content map predicted by Random Forest.

3.7. Prediction of electrical conductivity (EC) using Random Forest

Electrical conductivity was predicted with strong accuracy (R^2 = 78.7%, RMSE = 0.019 dS m^{-1}, MAE = 0.013 dS m^{-1}), indicating reliable model performance. This level of accuracy is consistent with previous digital soil mapping studies [41, 43, 62, 63, 64], which report that EC is often better predicted than other dynamic soil properties like SOC due to its sensitivity to soil texture, moisture status, and exchangeable cation content.

Important predictors (**Fig. 15**) for EC included silt, Na, DEM, TWI, slope, drainage density, LULC, and spectral indices such as BSI and RI, in much alignment with findings by Kaya et al. [60], Guan et al. [62], Park et al. [63], Amare et al. [47] and Suleymanov et al. [43]. From a pedological perspective, the strong influence of silt on exchangeable bases such as Na aligns with established relationships, as finer particles and higher base saturation increase ionic concentration in the soil solution [64]. The correlation analysis confirmed strong positive associations between EC and silt, Ca, Na, and K, and a negative association with sand content. The importance of terrain attributes further reflects moisture accumulation in lower landscape positions [47], which enhances salt retention.

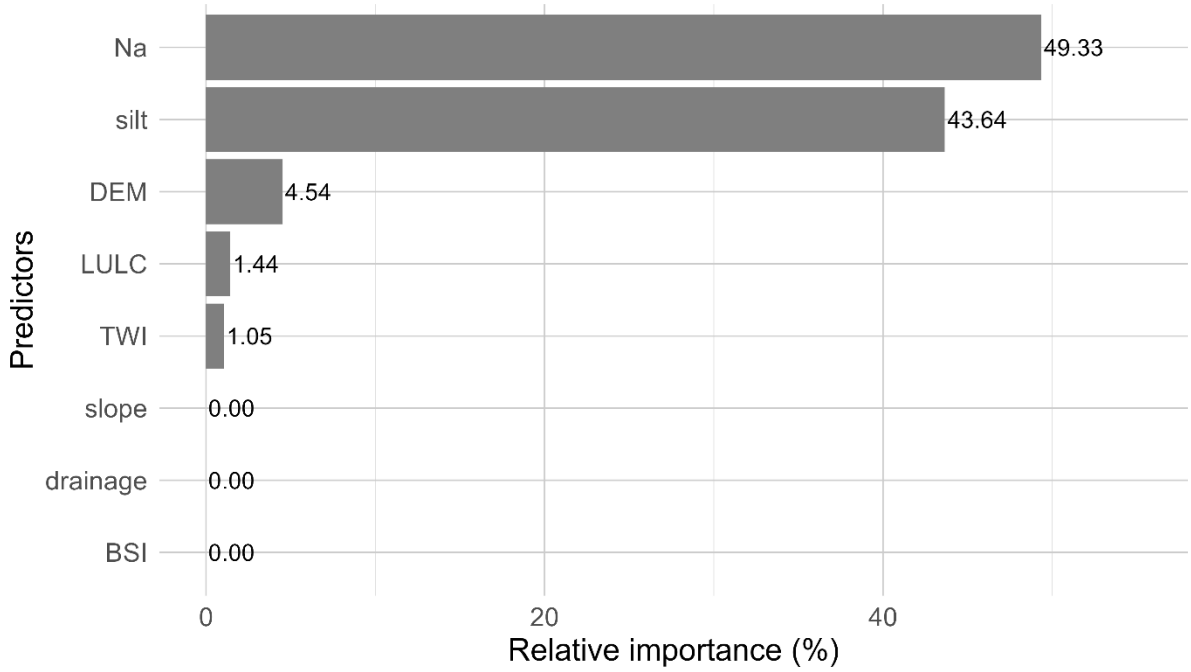


Fig. 15. The relative importance of environmental variables for predicting EC.

Predicted EC values (**Table 5**) predominantly fall within non-saline to slightly saline categories [56], which are generally suitable for tea cultivation. Tea is known to be

sensitive to salinity, and EC values below critical thresholds are favourable for root development and nutrient uptake. The RF predicted EC map (**Fig. 16**) therefore provides a robust basis for evaluating salinity-related constraints in subsequent tea suitability assessments within the study area.

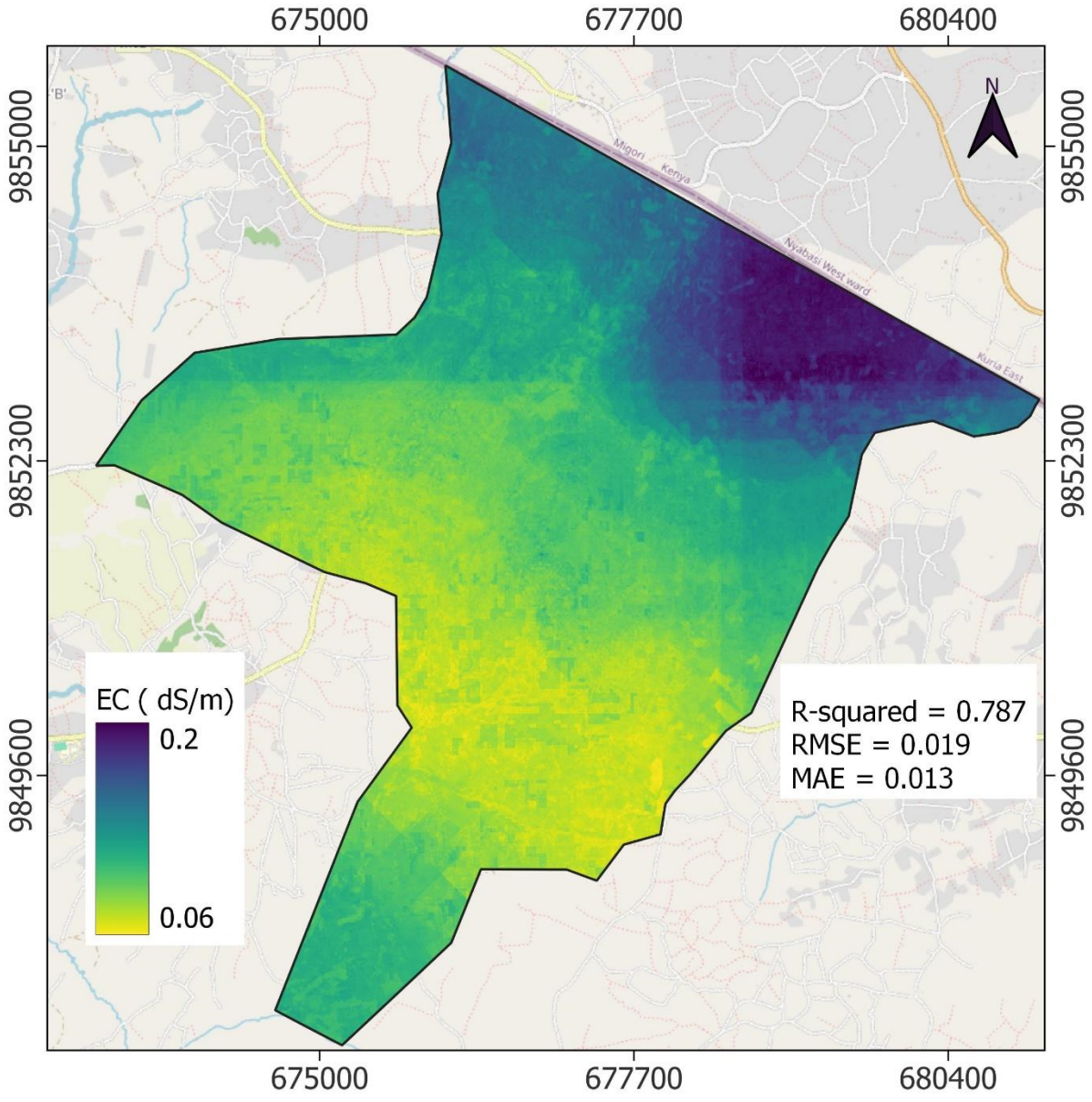


Fig. 16. Prediction map of EC produced by Random Forest

Table 5. Prediction summary statistics.

		Prediction summary statistics			
Soil property	**Unit**	**min**	**max**	**mean**	**sd**
Silt	*%*	17.41	35.99	23.88	4.38
EC	*dS m⁻¹*	0.06	0.21	0.11	0.03
Soil pH		4.25	4.77	4.54	0.12
SOC	*%*	2.1	4.86	3.6	0.57
K	*cmol(+) kg⁻¹*	0.39	1.59	0.79	0.22
P	*mgkg⁻¹*	2.80	8.74	4.27	0.89
Clay	*%*	18	35	26	3

pH=soil pH, EC=electrical conductivity, P = available phosphorus, SOC = soil organic carbon, K = exchangeable potassium

4. Conclusion

The present study predicted key soil properties in the selected tea potential area using Random Forest (RF) machine learning approach within a Digital Soil Mapping (DSM) framework, integrating SCORPAN-based environmental covariates which included DEM-derived topographic variables, spectral indices from Sentinel-2A, and spatial soil information. Model performance for properties with positive predictive power in terms of coefficient of determination (R^2) was within the range of 36.4%—78.7%. The RF model most accurately predicted electrical conductivity (EC) (R^2 = 78.7%, RMSE = 0.019 dS m^{-1}, MAE = 0.013 dS m^{-1}), which is highly sensitive to soil texture and moisture status. This was followed by silt content (R^2 = 67.1%, RMSE = 0.124%, MAE = 0.090%), likely due to its strong linkage with topographic redistribution and moisture dynamics captured effectively by topographic variables and spectral indices. Prediction of soil reaction (pH) demonstrated a good accuracy (R^2 = 57.4%, RMSE = 0.018, MAE = 0.018) as the model successfully captured texture-mediated buffering and topography-modulated leaching processes. Exchangeable Potassium (K) (R^2 = 49.0%, RMSE = 0.135 cmol (+) kg^{-1}, MAE = 0.095 cmol (+) kg^{-1}) was predicted with moderate accuracy. Also, soil organic carbon (SOC) (R^2 = 44.7%, RMSE = 0.155%, MAE = 0.101%), and available phosphorus (P) (R^2 =

0.409, RMSE = 0.257 mgkg^{-1}, MAE = 0.188 mg kg^{-1}) showed moderate performance, reflecting the high small-scale variability introduced by intensive cultivation and masking effects of localized management such as fertilizer inputs. Clay content (R^2 = 0.364, RMSE = 0.197%, MAE = 0.132%) yielded the lowest positive accuracies attributed to complex mineralogical variability respectively. Generally, the disparity in model prediction results of this study reflected the influence of intensive cultivation and recurrent anthropogenic inputs, such as fertilizer application, which obscure inherent soil–landscape relationships and introduce fine-scale variability that is not adequately captured by broader environmental covariates.

Predicted soil property maps indicate a moderately suitable fertility status for tea, characterized by acidic to moderately acidic environments (pH 4.2–4.8) and non-saline EC levels. However, low to moderate levels of SOC, K, and P suggest that nutrient deficiencies could constrain productivity. Generally, the results indicate that the study area is largely suitable for tea cultivation, particularly regarding soil reaction and texture. To improve tea productivity, efforts should focus on organic matter management and site-specific nutrient management to address P and K deficiencies. These findings serve as a vital tool for evidence-based decision-making and farm-level planning. Nevertheless, future studies are essential to overcome limitations such as limited sampling density and the need for additional covariates that capture management history. Furthermore, future research should focus on hybrid approaches, such as Random Forest combined with residual kriging, to better capture unexplained spatial dependence in highly managed landscapes.

Funding

Funder: Higher Education for Economic Transformation (HEET) project.

Project Number: P166415

Author contributions statement

FFM conceptualized the study, curated the data, conducted the formal analysis and investigation, developed the methodology, acquired funding and resources, performed the software-based analysis and validation, prepared the visualizations, and wrote the original manuscript draft. **BHJM** and **HGS** contributed to the study conceptualization, data curation, formal analysis, investigation, methodology, supervision, and validation. **BEM** and **EF** contributed to the conceptualization, data curation, investigation, methodology, and validation. All authors reviewed and edited the manuscript and approved the final version.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics, Consent to Participate, and Consent to Publish

Not applicable.

Data availability

The data used in this study is published in the Mendeley Data repository at https://doi.org/10.17632/NXXTTCHHJW

Acknowledgements

We acknowledge a financial support from the Higher Education for Economic Transformation (HEET) project, implemented under Tanzania's Ministry of Education, Science and Technology. We also thank the Tarime District Council and Ganyange Ward administration for granting permission to conduct this study in Ganyange Ward. Special appreciation is extended to the Agricultural Officer of Ganyange Ward and his team for their valuable assistance during field data collection.

References

1. Adeniyi OD, Bature H, Mearker M. A Systematic Review on Digital Soil Mapping Approaches in Lowland Areas. Land. 2024;13(3):379. https://doi.org/10.3390/land13030379
2. Munghemezulu C, Kganyago M, Mashaba-Munghemezulu Z, Chirima GJ, Weepener H, Nduku L. Regional-scale digital soil mapping of key GlobalSoilMap properties using machine learning and MODIS data in South Africa. Soil Advances. 2025; 4:100078. https://doi.org/10.1016/j.soilad.2025.100078
3. Mondal BP, Sahoo RN, Das B, Chakraborty D, Kumar N, Raj GB. Digital Soil Mapping: concepts, methods, and applications - Remote sensing and GIS perspectives. In Remote Sensing of Soils: Mapping, Monitoring, and Measurement (pp. 195–212). 2023. https://doi.org/10.1016/B978-0-443-18773-5.00018-1
4. Nenkam AM, Wadoux AMJC, Minasny B, Silatsa FBT, Yemefack M, Ugbaje SU, Akpa S, Zijl GVan, Bouasria A, Bouslihim Y, Chabala LM, Ali A, McBratney AB. Applications and challenges of digital soil mapping in Africa. Geoderma, 449. 2024. https://doi.org/10.1016/j.geoderma.2024.117007

5. Hengl T, Heuvelink GBM, Kempen B, Leenaars JGB, Walsh MG, Shepherd KD, Sila A, MacMillan RA, De Jesus JM, Tamene L, Tondoh JE. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PLoS ONE, 10(6). 2015. https://doi.org/10.1371/journal.pone.0125814

6. Wadoux AMJC, Minasny B, McBratney AB. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. In Earth-Science Reviews (Vol. 210). 2020. https://doi.org/10.1016/j.earscirev.2020.103359

7. Takoutsing B. Digital soil mapping using uncertain soil observations to support agricultural intensification in West and Central Africa [Wageningen University]. 2024. https://search.proquest.com/openview/faf7ad3e76d632fcd9a0bae1a4b47bb4/1?pq-origsite=gscholar&cbl=2026366&diss=y

8. Kempen B, Brus DJ, Stoorvogel JJ, Heuvelink GBM, de Vries F. Efficiency Comparison of Conventional and Digital Soil Mapping for Updating Soil Maps. Soil Science Society of America Journal. 2012;76(6):2097–2115. https://doi.org/10.2136/sssaj2011.0424

9. Adeniyi OD, Brenning A, Bernini A, Brenna S, Maerker M. Digital Mapping of Soil Properties Using Ensemble Machine Learning Approaches in an Agricultural Lowland Area of Lombardy, Italy. Land, 12(2). 2023. https://doi.org/10.3390/land12020494

10. Mgohele R, Massawe BHJ, Shitindi M, Sanga HG, Omar MM. Land suitability assessment for sisal production: A machine learning and Analytical Hierarchy Process integrated approach. Soil Advances. 2025; 3:100048. https://doi.org/10.1016/j.soilad.2025.100048

11. Behrens T, Schmidt K, MacMillan RA, Viscarra Rossel RA. Multi-scale digital soil mapping with deep learning. Scientific Reports, 8(1). 2018.

https://doi.org/10.1038/s41598-018-33516-6

12. Massawe BHJ. Digital Soil Mapping and GIS-based Land Evaluation for Rice Suitability in Kilombero [The Ohio State University]. 2015. https://search.proquest.com/openview/666013e1fdc137d99bda5ca960df793e/1?pq-origsite=gscholar&cbl=18750
13. Massawe BHJ, Slater BK, Subburayalu SK, Kaaya AK, Winowiecki L. Updating legacy soil maps for climate resilient agriculture: A case of kilombero valley, Tanzania. In Climate Change and Multi-Dimensional Sustainability in African Agriculture: Climate Change and Sustainability in Agriculture (pp. 345–364). Springer International Publishing. 2016. https://doi.org/10.1007/978-3-319-41238-2_19
14. Kempen B, Dalsgaard S, Kaaya AK, Chamuya N, Ruipérez-González M, Pekkarinen A, Walsh MG. Mapping topsoil organic carbon concentrations and stocks for Tanzania. Geoderma. 2019; 337:164–180. https://doi.org/10.1016/j.geoderma.2018.09.011
15. Kimaro DN, Msanya BM, ArakI S, Otsuka H. Application of computerized land evaluation systems in Tanzania: a case study in Kilosa District, Morogoro Region. UNISWA Journal of Agriculture. 2004;10(1):40–50. https://doi.org/10.4314/uniswa.v10i1.4606
16. Msanya B, Magoggo J, Otsuka H. Development of Soil Surveys in Tanzania. Pedologist, VOL.46, 79–88. 2002. https://www.jstage.jst.go.jp/article/pedologist/46/2/46_KJ00007984927/_article/-char/ja/
17. Magoggo JP. Considerations for an Integrated Land Evaluation System for Tanzania. Lessons from the Land Suitability Assessment of Mbulu District. In National Service, TARI Mlingano. 1998. https://wgbis.ces.iisc.ac.in/energy/HC270799/LM/SUSLUP/Thema3/190/190.pdf

18. Massawe BHJ, Subburayalu SK, Kaaya AK, Winowiecki L, Slater BK. Mapping numerically classified soil taxa in Kilombero Valley, Tanzania using machine learning. Geoderma. 2018; 311:143–148. https://doi.org/10.1016/j.geoderma.2016.11.020
19. Maro GP, Mbwambo SG, Monyo HE, Mosi EJ. Use of Remote Sensing and GIS Techniques for Land Use Mapping at Four Selected TaCRI Stations, Tanzania. Journal of Geosciences and Geomatics. 2024;12(1):6–11. https://doi.org/10.12691/jgg-12-1-2
20. Chen Fu, B., Xie Y, Bai S, Cui X. Integrating extreme temperature metrics into tea land suitability models: a GIS-based assessment in the ecological and cultural tourism circle in Western Hubei, China. Theoretical and Applied Climatology. 2025;156(5):276. https://doi.org/10.1007/s00704-025-05518-8
21. Tarime District Council. Tarime Strategic Plan from 2021 to 2026. 2021. https://www.google.com/search?q=Tarime+Strategic+Plan+from+2021+to+2026.+Tarime+District+Council+as+Mandated+by+The+United+Republic+of+Tanzania+President's+Office%2C+Regional+Administration%2C+and+Local+Government+through&oq=Tarime+Strategic+Plan+from+202
22. ISRIC. Soils of the World, according to the World Reference Base for Soil Resources (WRB 2014, Update 2015). ISRIC. 2015. https://data.isric.org/geonetwork/srv/eng/catalog.search#/metadata/wrb2014-map
23. Brus DJ. Sampling for digital soil mapping: A tutorial supported by R scripts. In Geoderma (Vol. 338, pp. 464–480). Elsevier. 2019. https://doi.org/10.1016/j.geoderma.2018.07.036
24. FAO. Guidelines for soil description. Rome.Italy. In Chest (Issue January). FAO. 2006. https://books.google.com/books/about/Guidelines_for_Soil_Description.html?id=x

0_PVec33LgC

25. Bouyoucos GJ. Hydrometer Method Improved for Making Particle Size Analyses of Soils 1. Agronomy Journal. 1962;54(5):464–465. https://doi.org/10.2134/agronj1962.00021962005400050028x

26. Simon B, Tolner L, Rèkási M, Michéli E. Soil acidity investigation by potentiometric titrations. Cereal Research Communications. 2006;34(1):283–286. https://doi.org/10.1556/CRC.34.2006.1.71

27. Sumner ME, Miller WP. Cation exchange capacity and exchange coefficients. In Methods of Soil Analysis, Part 3: Chemical Methods (pp. 1201–1229). wiley. 2018. https://doi.org/10.2136/sssabookser5.3.c40

28. Walkley A, Black IA. An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Science. 1934;37(1):29–38. https://doi.org/10.1097/00010694-193401000-00003

29. Bray RH, Kurtz LT. Determination of total, organic, and available forms of phosphorus in soils. Soil Science. 1945;59(1):39–45. https://doi.org/10.1097/00010694-194501000-00006

30. Olsen SR, Cole CV, Watandbe F, Dean L. Estimation of Available Phosphorus in Soil by Extraction with sodium Bicarbonate. Journal of Chemical Information and Modeling. 1954;53(9):1689–1699. https://books.google.com/books?hl=en&lr=&id=d-oaM88x5agC&oi=fnd&pg=PA2&dq=Olsen,+S.R.,+1954.+Estimation+of+available+phosphorus+in+soils+by+extraction+with+sodium+bicarbonate+(Issue+939).+US+Department+of+Agriculture&ots=z_-mXGkU_B&sig=iaFZyCN3ilvpcrTuUu9

31. Jackson MLR. Soil Chemical Analysis. M.L. Jackson. 1967. https://books.google.co.tz/books?id=V38XAQAAIAAJ

32. Rhoades JD. Soluble Salts (pp. 167–179). John Wiley & Sons, Ltd. 1982. https://doi.org/10.2134/agronmonogr9.2.2ed.c10

33. Bremner JM. Total nitrogen. In Methods of Soil Analysis, Part 2: Chemical and Microbiological Properties (pp. 1149–1178). John Wiley & Sons, Ltd. 2016. https://doi.org/10.2134/agronmonogr9.2.c32

34. Pahlevan N, Sarkar S, Franz BA, Balasubramanian SV, He J. Sentinel-2 MultiSpectral Instrument (MSI) data processing for aquatic science applications: Demonstrations and validations. Remote Sensing of Environment. 2017;201(11):47–56. https://doi.org/10.1016/j.rse.2017.08.033

35. Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L, Seal D, Shaffer S, Shimada J, Umland J, Werner M, Oskin M, Burbank D, Alsdorf DE. The shuttle radar topography mission. Reviews of Geophysics. 2007;45(2):2004. https://doi.org/10.1029/2005RG000183

36. Malone BP, McBratney AB, Minasny B, Laslett GM. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma. 2009;154(1-2):138-152. https://doi.org/10.1016/j.geoderma.2009.10.007

37. Dai L, Wang Z, Zhuo Z, Ma Y, Shi Z, Chen S. Prediction of soil organic carbon fractions in tropical cropland using a regional visible and near-infrared spectral library and machine learning. Soil and Tillage Research, 245. 2025. https://doi.org/10.1016/j.still.2024.106297

38. Gomes LC, Faria RM, de Souza E, Veloso GV, Schaefer CEGR, Filho EIF. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma. 2019; 340:337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

39. Grimm R, Behrens T, Märker M, Elsenbeer H. Soil organic carbon concentrations and stocks on Barro Colorado Island - Digital soil mapping using Random Forests

analysis. Geoderma. 2008;146(1–2):102–113. https://doi.org/10.1016/j.geoderma.2008.05.008

40. Canero FM, Rodriguez-Galiano V, Aragones D. Machine Learning and Feature Selection for soil spectroscopy. An evaluation of Random Forest wrappers to predict soil organic matter, clay, and carbonates. Heliyon, 10(9). 2024. https://doi.org/10.1016/j.heliyon.2024.e30228

41. Dharumarajan S, Hegde R, Singh SK. Spatial prediction of major soil properties using Random Forest techniques - A case study in semi-arid tropics of South India. Geoderma Regional. 2017; 10:154–162. https://doi.org/10.1016/j.geodrs.2017.07.005

42. Naimi S, Ayoubi S, Demattê JAM, Zeraatpisheh M, Amorim MTA, Mello FAde O. Spatial prediction of soil surface properties in an arid region using synthetic soil image and machine learning. Geocarto International. 2022;37(25):8230–8253. https://doi.org/10.1080/10106049.2021.1996639

43. Suleymanov A, Gabbasova I, Komissarov M, Suleymanov R, Garipov T, Tuktarova I, Belan L. Random Forest Modeling of Soil Properties in Saline Semi-Arid Areas. Agriculture (Switzerland), 13(5). 2023. https://doi.org/10.3390/agriculture13050976

44. Tamiru B, Soromessa T, Warkineh B, Legese G. Modelling selected soil chemical properties using TerrSet and Random Forest: a case of Hangadi watershed, Oromia, Ethiopia. In Geology, Ecology, and Landscapes (Vol. 9, Issue 4, pp. 1177–1193). Taylor and Francis Ltd. 2023. https://doi.org/10.1080/24749508.2024.2429842

45. Were K, Bui DT, Dick, Ø. B., Singh BR. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecological

Indicators. 2015; 52:394–403. https://doi.org/10.1016/j.ecolind.2014.12.028

46. Zhang H, Wu P, Yin A, Yang X, Zhang M, Gao C. Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison of multiple linear regressions and the random forest model. Science of the Total Environment. 2017; 592:704–713. https://doi.org/10.1016/j.scitotenv.2017.02.146
47. Amare H, Admase H, Ewunetu T. Influence of land-use types and topographic slopes on the physico-chemical characteristics of soils in Northwestern Ethiopia. Frontiers in Soil Science, 4. 2024. https://doi.org/10.3389/fsoil.2024.1463315
48. Zhang Qu, X., Song X, Xiao Y, Wang A, Li D. Spatial Variation in Soil Base Saturation and Exchangeable Cations in Tropical and Subtropical China. Agronomy, 13(3). 2023. https://doi.org/10.3390/agronomy13030781
49. Landon JR. Booker Tropical Soil Manual: A Handbook for Soil Survey and Agricultural Land Evaluation in the Tropics and Subtropics. In Booker Tropical Soil Manual: A Handbook for Soil Survey and Agricultural Land Evaluation in the Tropics and Subtropics. Taylor and Francis. 2014. https://doi.org/10.4324/9781315846842
50. Xing W, Zhou C, Li J, Wang W, He J, Tu Y, Cao X, Zhang Y. Suitability Evaluation of Tea Cultivation Using Machine Learning Technique at Town and Village Scales. Agronomy. 2022;12(9):2010. https://doi.org/10.3390/agronomy12092010
51. Makungwe M, Chabala LM, Chishala BH, Lark RM. Performance of linear mixed models and random forests for spatial prediction of soil pH. Geoderma, 397. 2021. https://doi.org/10.1016/j.geoderma.2021.115079
52. Abdule AM, Woyesa AK. Physical Land Suitability Assessment for Tea Cultivation Using GIS-Based Multi-Criteria Approach at Dabus Basin of Oromia Region, Southwest Ethiopia. American Journal of Agriculture and Forestry. 2023;11(2):45–57. https://doi.org/10.11648/j.ajaf.20231102.12

53. Chen P, Li C, Chen S, Li Z, Zhang H, Zhao C. Tea Cultivation Suitability Evaluation and Driving Force Analysis Based on AHP and Geodetector Results: A Case Study of Yingde in Guangdong, China. Remote Sensing, 14(10). 2022. https://doi.org/10.3390/rs14102412

54. Das AC, Noguchi R, Ahamed T. Integrating an expert system, gis, and satellite remote sensing to evaluate land suitability for sustainable tea production in bangladesh. Remote Sensing. 2020;12(24):1–25. https://doi.org/10.3390/rs12244136

55. Das AC, Noguchi R, Ahamed T. Integrating Analytical Hierarchy Process with GIS and Satellite Remote Sensing to Assess Land Suitability for Sustainable Tea Production in Bangladesh. In Springer (pp. 205–237). 2024. https://doi.org/10.1007/978-981-97-1188-8_8

56. Gahlod NS, Binjola S, Ravi R, Arya VS. Land-site suitability evaluation for tea, cardamom and rubber using Geo-spatial technology in Wayanad district, Kerala. Journal of Applied and Natural Science. 2017;9(3):1440–1447. https://doi.org/10.31018/jans.v9i3.1381

57. Jayasinghe SL, Kumar L, Sandamali J. Assessment of potential land suitability for tea (Camellia sinensis (L.) O. Kuntze) in Sri Lanka using a gis-based multi-criteria approach. Agriculture (Switzerland), 9(7). 2019. https://doi.org/10.3390/agriculture9070148

58. Mallick M, Krishnaiah YV, Panja K, Das D, Rai D, Hati M, Chakma A. Land suitability assessment for tea cultivation in Jalpaiguri district of West Bengal, India, using AHP and DEMATEL techniques. Environment, Development and Sustainability. 2024. https://doi.org/10.1007/s10668-024-05711-1

59. Isong IA, John K, Okon PB, Ogban PI, Afu SM. Soil quality estimation using environmental covariates and predictive models: an example from tropical soils of

Nigeria. Ecological Processes. 2022;11(1):66. https://doi.org/10.1186/s13717-022-00411-y

60. Kaya F, Schillaci C, Keshavarzi A, Başayiğit, L. Predictive Mapping of Electrical Conductivity and Assessment of Soil Salinity in a Western Türkiye Alluvial Plain. Land, 11(12). 2022. https://doi.org/10.3390/land11122148

61. Mgohele Massawe, B. H. J., Shitindi MJ, Sanga HG, Omar MM. Prediction of soil texture using remote sensing data. A systematic review. In Frontiers in Remote Sensing (Vol. 5). Frontiers Media SA. 2024. https://doi.org/10.3389/frsen.2024.1461537

62. Guan Y, Grote K, Schott J, Leverett K. Prediction of Soil Water Content and Electrical Conductivity using Random Forest Methods with UAV Multispectral and Ground-coupled Geophysical Data. Remote Sensing, 14(4). 2022. https://doi.org/10.3390/rs14041023

63. Park HJ, Baek N, Seo BS, Jeong YJ, Yang HI, Lee SI, Yoon KS, Kim HY, Choi WJ. Estimation of the electrical conductivity of saturated paste from soil–water extracts of coastal saline paddy soils using random forest and multiple regression models. Journal of Soils and Sediments. 2024;24(3):1250–1259. https://doi.org/10.1007/s11368-024-03725-w

64. Win K, Sato T, Hiroshima T. Applicability of remote sensing and machine learning for predicting bulk soil electrical conductivity under different forest types in central Japan. Soil Advances. 2025;3(5):100045. https://doi.org/10.1016/j.soilad.2025.100045