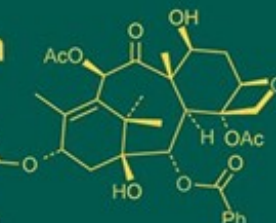
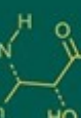
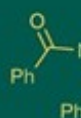
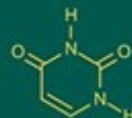


International Journal of Advanced Biochemistry Research



ISSN Print: 2617-4693
 ISSN Online: 2617-4707
 NAAS Rating (2026): 5.29
 Indexing: Embase
 IJABR 2026; SP-10(10): 522-527
www.biochemjournal.com
 Received: 16-08-2026
 Accepted: 26-09-2026
 Published: 07-10-2026

The names and affiliations of all authors are provided below, following the references.

Digital mapping of available phosphorus in agricultural soils of Tamil Nadu using Random Forest and environmental covariates

B Bhanukiran Reddy¹, S Maragatham², R Santhi³, D Balachandar⁴, D Vijaya Lakshmi⁵, V Davamani⁶, D Vasu⁷, M Gopalakrishnan⁸

DOI: <https://www.doi.org/10.33545/26174693.2026.v10.i10Sg.10333>

ABSTRACT

Spatial information on available phosphorus (P) is important for identifying areas of contrasting soil fertility and for supporting location-specific nutrient management. This study evaluated the spatial variability of available P in agricultural soils of Tamil Nadu, India, and predicted its distribution using Random Forest (RF) modelling and environmental covariates. A total of 250 georeferenced surface soil samples were collected from 0-15 cm depth using conditioned Latin Hypercube Sampling to represent the major agro-ecological and soil conditions of the state. Available P was determined using Bray-1 extraction for acidic soils and Olsen extraction for neutral and alkaline soils. Spectral and terrain covariates derived from Landsat 8 and Shuttle Radar Topography Mission data were screened using the *Boruta* feature-selection algorithm and subsequently used for RF modelling. Available P ranged from 4.0 to 102.0 kg ha⁻¹, with a mean of 39.3 kg ha⁻¹. *Boruta* identified slope, normalized difference vegetation index and green reflectance as the principal predictors of P variability. The optimized RF model, with mtry = 3 and ntree = 1000, produced an R^2 of 0.75, RMSE of 11.56 kg ha⁻¹ and MAE of 9.56 kg ha⁻¹. Spatial prediction indicated comparatively higher available P in parts of southern and central Tamil Nadu, whereas lower values occurred in portions of northern and western Tamil Nadu. The results show that integrating georeferenced soil observations with remotely sensed and terrain covariates can delineate regional-scale variation in available P and provide a spatial basis for targeted soil testing and phosphorus management.

Keywords: Available phosphorus, *Boruta*, Digital soil mapping, Landsat 8, NDVI, Random Forest, Soil fertility, Tamil Nadu

1. Introduction

Phosphorus (P) is a major plant nutrient involved in energy transfer, root development, nucleic acid synthesis and several metabolic processes that determine crop establishment and productivity. Although agricultural soils may contain appreciable quantities of total P, only a relatively small fraction is available for immediate plant uptake because phosphorus is readily retained by soil constituents. Its availability consequently varies with soil reaction, mineral composition, parent material, weathering environment and management history.

The agricultural soils of Tamil Nadu occur under contrasting climatic, physiographic and edaphic conditions. Red, black, alluvial and lateritic soils are widely distributed across the state and are represented mainly by Alfisols, Vertisols, Inceptisols, Entisols and Ultisols. These soils differ considerably in texture, mineralogy, calcium carbonate content, soil reaction and nutrient-retention behaviour. Such heterogeneity is expected to produce pronounced spatial variation in available P and limits the usefulness of generalized nutrient assumptions for large regions.

Phosphorus availability is particularly sensitive to soil reaction. Under acidic conditions, P may be retained by iron- and aluminium-bearing soil constituents, whereas in alkaline and calcareous soils it may be converted into sparingly soluble calcium-associated forms. In addition, fertilizer history, crop uptake, erosion, deposition and redistribution of surface soil can alter the spatial pattern of P availability. Consequently, the P status measured at individual sampling locations cannot by itself describe the continuous regional distribution needed for spatially targeted nutrient management.

Digital soil mapping provides a framework for extending point observations across landscapes by relating measured soil attributes to spatially exhaustive environmental covariates. Terrain derivatives can represent processes such as runoff, erosion and deposition, while satellite-based spectral and vegetation indices provide information on land-surface characteristics and vegetation condition. Machine-learning approaches are

Corresponding Author:
 B Bhanukiran Reddy
 Tamil Nadu Agricultural
 University, Coimbatore 641
 003, Tamil Nadu, India

particularly useful in this context because relationships between soil properties and environmental controls are often nonlinear and interactive.

Random Forest (RF) is an ensemble-learning method that combines multiple decision trees and can accommodate complex relationships among predictors without requiring strict assumptions concerning variable distributions (Breiman, 2001) [1]. RF has been widely used in digital soil mapping because it can integrate large sets of environmental covariates and is relatively robust to nonlinearities and predictor interactions (Belgiu *et al.*, 2016; Bhunia *et al.*, 2018) [2, 3]. Feature-selection methods can further improve interpretability by reducing the predictor set to variables that carry useful information for a specific soil attribute.

Previous work from the broader soil survey used RF to assess soil pH and soil organic matter (Reddy *et al.*, 2024) [4]. The present study isolates available P as an independent nutrient endpoint and evaluates its regional spatial pattern using the existing survey and environmental-covariate framework. The objectives were to (i) characterize the variability of available P in the sampled agricultural soils, (ii) identify the principal environmental covariates associated with its spatial distribution, and (iii) generate a continuous prediction of available P across Tamil Nadu using RF modelling.

2. Materials and Methods

2.1 Study area

The study covered agricultural soils distributed across Tamil Nadu, southern India. The region includes coastal plains, river valleys, semi-arid interior landscapes and hilly terrain associated with the Western Ghats. These physiographic differences are accompanied by substantial variation in rainfall, land use, soil-forming materials and agricultural production systems. The sampled soils represented the major soil orders occurring in the study area, including Alfisols, Inceptisols, Vertisols, Entisols and Ultisols.

2.2 Sampling design and soil collection

Sampling locations were selected using conditioned Latin Hypercube Sampling (cLHS) to represent the environmental variability of the study region. Variables considered during sampling-point selection included land use and land cover, agro-ecological information, soil sub-order, digital elevation information, terrain parameters such as slope and aspect, road-network data and normalized difference vegetation index. This procedure resulted in 250 sampling locations distributed across the study area.

Surface soil samples were collected from 0-15 cm depth. At each location, five subsamples were collected within approximately a 1-m radius and combined to obtain one composite sample. Each sampling point was georeferenced using GPS. Samples were transported to the laboratory, air-dried and passed through a 2-mm sieve before chemical analysis.

2.3 Determination of available phosphorus

Available phosphorus was determined using soil-reaction-specific standard extraction procedures. Bray-1 extraction was used for acidic soils following Bray *et al.* (1945) [5], while the Olsen method was used for neutral and alkaline soils following Olsen *et al.* (1954) [6]. Available P values are presented in kg ha⁻¹, following the units used in the thesis tables and mapping outputs.

2.4 Environmental covariates

Environmental covariates were derived from satellite imagery and terrain data. Spectral information was obtained from Landsat 8, while terrain attributes were derived from Shuttle Radar Topography Mission (SRTM) data. The broader covariate set included spectral bands and vegetation-, moisture- and soil-related indices together with terrain variables. Variables represented in the original covariate dataset included blue, green, red and near-infrared reflectance; NDVI; soil-adjusted vegetation index; transformed NDVI; ratio vegetation

index; difference vegetation index; slope; aspect; topographic wetness index; valley-bottom flatness; and ridge-top flatness.

2.5 Feature selection

The Boruta algorithm was used to identify environmental variables associated with available P before development of the final prediction model. Boruta creates randomized shadow variables from the original predictors and compares their importance with that of the observed predictors using Random Forest-based importance estimates. Predictors consistently performing better than the shadow variables are retained as important, whereas poorly performing predictors are rejected. For available P, the analysis retained slope, NDVI and green reflectance as the principal covariates.

2.6 Random Forest modelling and performance assessment

Random Forest regression was used to predict available P from the selected environmental covariates. Model tuning evaluated the number of variables considered at each split (mtry) and the number of trees (ntree). The tuning output in the original analysis used repeated cross-validation as the optimization criterion. The best-performing configuration for available P was mtry = 3 and ntree = 1000. Model performance was summarized using the coefficient of determination (R²), root mean squared error (RMSE) and mean absolute error (MAE). The optimized RF model was subsequently applied to the environmental-covariate layers to generate the continuous spatial prediction of available P across Tamil Nadu.

$$RMSE = \sqrt{[(1/n) \sum (y_i - \hat{y}_i)^2]}$$

$$MAE = (1/n) \sum |y_i - \hat{y}_i|$$

$$R^2 = 1 - [\sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2]$$

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

3. Results

3.1 Descriptive variability of available phosphorus

Available P showed marked variation among the 250 agricultural soil samples. Values ranged from 4.0 to 102.0 kg ha⁻¹, with a mean of 39.3 kg ha⁻¹ and a median of 35.5 kg ha⁻¹. The standard deviation was 23.3 kg ha⁻¹ and the standard error was 1.68 kg ha⁻¹. The distribution showed a skewness of 0.37 and a kurtosis of -0.90, indicating broad variation across the sampled landscape without pronounced positive skewness (Table 1).

3.2 Selection of environmental predictors

Boruta feature selection identified slope, NDVI and green reflectance as the most important environmental variables for predicting available P. The selected predictors represent complementary aspects of the soil-landscape system: slope reflects terrain-related redistribution processes, whereas NDVI and green reflectance provide information on vegetation and land-surface characteristics. These variables were retained for RF modelling.

3.3 Random Forest model performance

The optimized RF model showed good agreement between observed and predicted available P. The model achieved an R² of 0.75, with an RMSE of 11.56 kg ha⁻¹ and an MAE of 9.56 kg ha⁻¹. The optimal configuration consisted of mtry = 3 and ntree = 1000 (Table 1). Increasing mtry beyond the selected value did not improve the model sufficiently to justify a more complex configuration.

The observed-versus-predicted relationship indicated that the model captured most of the regional variation in available P, although some high-P observations were underestimated. The tendency to compress extreme values was most evident at the upper end of the measured P range (Fig. 2c).

3.4 Spatial distribution of predicted available phosphorus

The RF prediction revealed clear spatial heterogeneity in available P across Tamil Nadu. Relatively higher predicted concentrations occurred in parts of the southern and central regions, whereas lower values were evident in portions of the northern and western regions (Fig.

2d). The resulting surface therefore delineated broad regional zones differing in predicted P availability rather than indicating a uniform nutrient status across the state.

4. Discussion

4.1 Regional variability in available phosphorus

The wide range in available P, from 4.0 to 102.0 kg ha⁻¹, demonstrates the heterogeneous P fertility of the sampled agricultural soils. This variation is consistent with the contrasting soil-forming environments represented in Tamil Nadu, where soil orders differ in texture, mineralogy, calcium carbonate content and reaction. Regional differences in cropping and nutrient management are also likely to contribute to the observed variability.

Phosphorus availability is governed by several competing processes. In acidic soils, reactions with Fe³⁺ and Al-bearing constituents can reduce soluble P, whereas in calcareous and alkaline soils P may be retained in calcium-associated forms. These mechanisms, together with crop removal and fertilizer addition, generate spatial variability that cannot be captured by a single generalized fertility value. The observed distribution therefore supports the need for spatially explicit assessment of P status.

4.2 Role of terrain in phosphorus distribution

Slope was one of the principal variables retained by *Boruta*. Although P is relatively immobile in soil solution, a considerable proportion of soil P is associated with the solid phase and can be redistributed through erosion and sediment transport. Sloping positions may lose nutrient-rich surface material, whereas lower or depositional positions may accumulate eroded particles. Terrain variables are therefore useful proxies for landscape processes that influence the spatial organization of soil nutrients.

The importance of slope in the RF model is consistent with the broader digital soil mapping literature, in which topographic attributes frequently contribute to predictions of soil properties in heterogeneous landscapes (Bhunia *et al.*, 2018) [3]. The result does not imply a single monotonic relationship between slope and available P; rather, RF can accommodate interactions between slope and other environmental variables across different landscape settings.

4.3 Information contributed by NDVI and green reflectance

NDVI was also selected as an important predictor. Vegetation indices are not direct measurements of soil phosphorus, but they integrate information related to canopy development, vegetation cover and land-surface conditions. These characteristics can co-vary with soil fertility, moisture regime, crop management and erosion protection. As a result, NDVI can contribute useful indirect information to regional nutrient prediction.

Green reflectance provided additional spectral information and was retained alongside NDVI. Its importance likely reflects differences in vegetation and exposed-surface characteristics across the agricultural landscape. The simultaneous selection of terrain and spectral variables emphasizes that available P at regional scale reflects a combination of landscape position and land-surface conditions rather than a single environmental control.

4.4 Predictive performance and model uncertainty

The RF model explained 75% of the observed variation in available P. Considering the broad environmental heterogeneity represented by the 250 sampling locations, this level of performance indicates that the selected environmental covariates contained substantial information relevant to regional P distribution. The RMSE of 11.56 kg ha⁻¹ and MAE of 9.56 kg ha⁻¹ provide a practical indication of prediction error relative to the observed range of 4–102 kg ha⁻¹.

The model showed greater disagreement at the upper end of the available P distribution. Random Forest models commonly predict extreme observations less accurately because ensemble predictions tend to move toward the centre of the response distribution. In the present dataset, localized factors such as fertilizer history or field-specific management may also contribute to high P values but were not represented explicitly in the environmental-covariate layers. Accordingly, the map should be interpreted as a regional prediction surface rather than as a field-scale fertilizer prescription.

4.5 Spatial pattern and implications for soil testing

Higher predicted available P occurred mainly in parts of southern and central Tamil Nadu, while lower predicted values were observed in portions of northern and western Tamil Nadu. The thesis discussion associated some higher-P areas with intensive agricultural activity and possible repeated P inputs. Because historical fertilizer-use data were not included as direct predictors, this explanation is best regarded as a plausible interpretation rather than a tested causal relationship.

The practical value of the prediction lies in its ability to identify broad zones where additional soil testing can be prioritized. Areas predicted to have lower P can be targeted for confirmatory sampling before fertilizer recommendations are made, whereas areas predicted to have comparatively high P can be examined to avoid unnecessary fertilizer application. Such stratification can improve the efficiency of soil-sampling programmes and support more spatially targeted nutrient management.

The prediction surface does not replace conventional soil testing. Instead, it complements point-based analyses by providing a regional context for where contrasting P conditions are likely to occur. Field-level recommendations should continue to rely on locally collected soil samples, crop requirement and established fertilizer recommendation procedures.

4.6 Scope and limitations

This study was intentionally restricted to the spatial prediction of available P using the environmental and modelling outputs available from the original soil survey. The approach captured regional variability but did not explicitly include fertilizer history, crop-specific nutrient inputs or other management records. In addition, the current analysis represents a regional-scale model based on the available Landsat 8 and SRTM covariates. These limitations should be considered when interpreting local predictions.

Future work could evaluate higher-resolution remote-sensing data, management covariates and independent spatial validation. Nevertheless, the present results demonstrate that the existing 250-point soil survey and environmental-covariate framework are sufficient to identify meaningful regional patterns in available P.

5. Conclusions

Available phosphorus varied substantially across the agricultural soils evaluated in Tamil Nadu, ranging from 4.0 to 102.0 kg ha⁻¹ with a mean of 39.3 kg ha⁻¹. *Boruta* feature selection identified slope, NDVI and green reflectance as the principal environmental covariates for predicting its spatial variation. The optimized Random Forest model, with *mtry* = 3 and *ntree* = 1000, achieved an *R*² of 0.75, RMSE of 11.56 kg ha⁻¹ and MAE of 9.56 kg ha⁻¹. The resulting prediction indicated comparatively higher available P in parts of southern and central Tamil Nadu and lower values in portions of northern and western Tamil Nadu. These findings show that georeferenced soil observations combined with remotely sensed and terrain-derived covariates can provide a useful regional representation of P availability. The generated map can be used to prioritize soil testing and to support spatially targeted phosphorus management, while field-specific fertilizer decisions should continue to be based on confirmatory soil analysis.

Acknowledgements

The authors acknowledge the institutional facilities and technical support provided for soil sampling, laboratory analysis, remote-sensing data processing and spatial modelling.

Conflict of interest

The authors declare no conflict of interest.

Data availability

The data used in this study were generated during the soil survey, laboratory analysis and spatial modelling undertaken as part of the doctoral research programme.

Tables

Table 1. Descriptive statistics and Random Forest model performance for available phosphorus in agricultural soils of Tamil Nadu.

Parameter	Value
Number of soil samples	250
Sampling depth	0-15 cm
Minimum available P	4.00 kg ha ⁻¹

Parameter	Value
Maximum available P	102.0 kg ha ⁻¹
Mean	39.3 kg ha ⁻¹
Median	35.5 kg ha ⁻¹
Mode	13.0 kg ha ⁻¹
Standard deviation	23.3 kg ha ⁻¹
Standard error	1.68 kg ha ⁻¹
Skewness	0.37
Kurtosis	-0.90
Selected RF predictors	Slope, NDVI, green reflectance
Optimal mtry	3
Optimal ntree	1000
R ²	0.75
RMSE	11.56 kg ha ⁻¹
MAE	9.56 kg ha ⁻¹

Note: Available P values and model metrics are reported as in the thesis tables and RF outputs. NDVI = normalized difference vegetation index; RMSE = root mean squared error; MAE = mean absolute error.

Figures

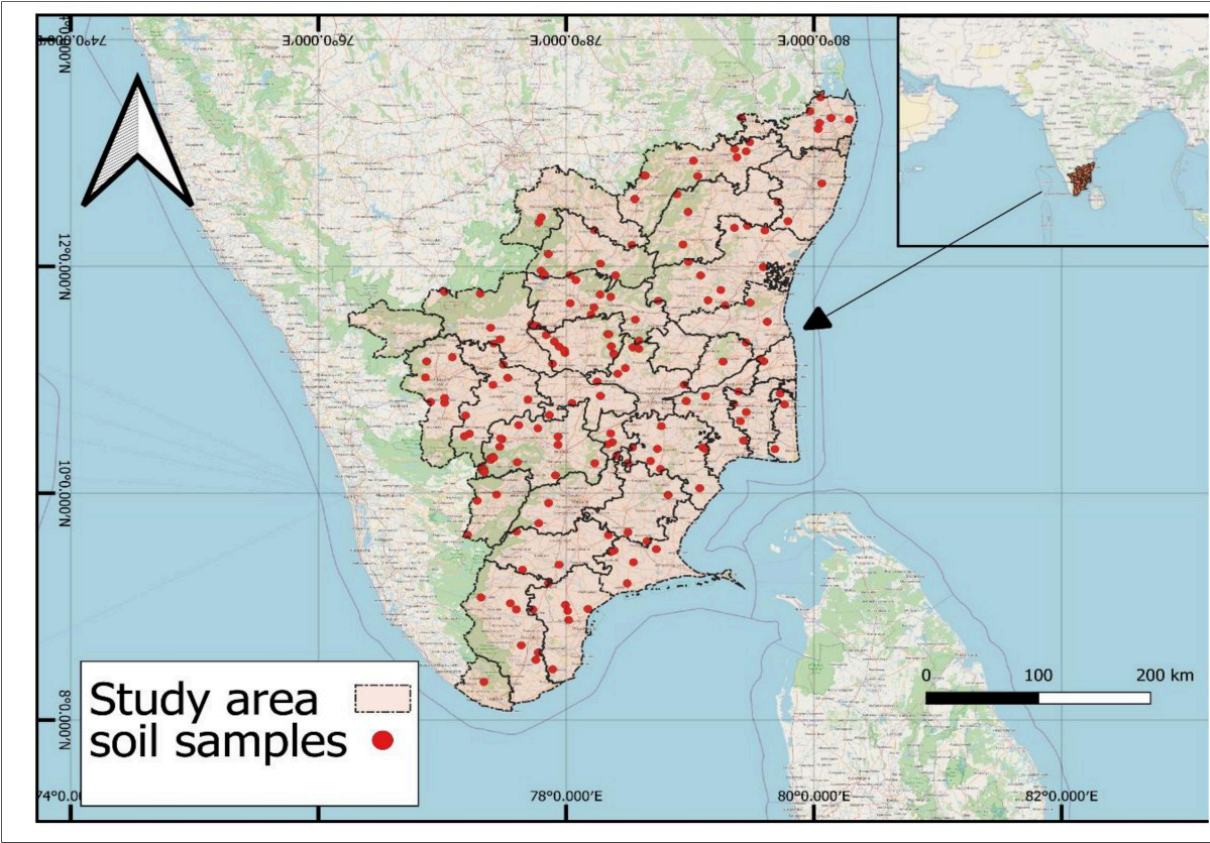


Figure 1. Study area and distribution of the georeferenced soil sampling locations across Tamil Nadu.

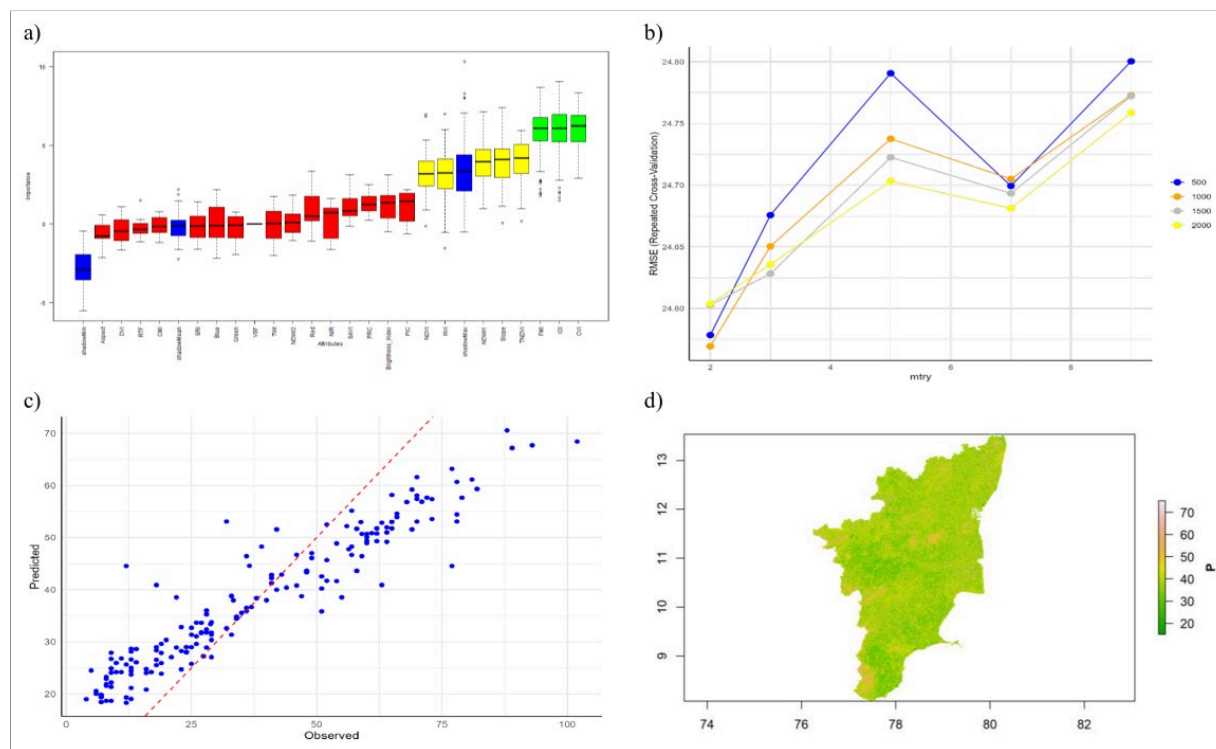


Figure 2. Random Forest analysis of available phosphorus distribution: (a) *Boruta* feature selection, (b) Random Forest tuning, (c) observed versus predicted available P (kg ha⁻¹), and (d) predicted spatial distribution of available P (kg ha⁻¹) across Tamil Nadu.

References

- Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32.
- Belgiu M, Drăguț L. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2016;114:24-31.
- Bhunia GS, Shit PK, Maiti R. Digital mapping of soil properties using Random Forests in a complex agro-ecological landscape. *Ecological Indicators*. 2018;93:493-503.
- Reddy BB, Maragatham S, Santhi R, Balachandrar D, Vijaya Lakshmi D, Davamani V, Vasu D, Gopalakrishnan M. Predictive soil mapping using random forest models: Applications in pH and soil organic matter assessment. *Plant Science Today*. 2024;11(4):463-474.
- Bray RH, Kurtz LT. Determination of total, organic, and available forms of phosphorus in soils. *Soil Science*. 1945;59:39-45.
- Olsen SR, Cole CV, Watanabe FS, Dean LA. Estimation of available phosphorus in soils by extraction with sodium bicarbonate. *USDA Circular 939*. United States Department of Agriculture, Washington, DC; 1954.
- Grimm R, Behrens T, Märker M, Elsenbeer H. Soil organic carbon concentrations and stocks on Barro Colorado Island: Digital soil mapping using Random Forests. *Geoderma*. 2008;146(1-2):102-113.
- Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002;2(3):18-22.
- Ma T, Brus DJ, Zhu AX, Zhang L, Scholten T. Comparison of conditioned Latin hypercube and feature space coverage sampling for predicting soil classes using simulation from soil maps. *Geoderma*. 2020;370:114366.
- Maragatham S, Santhi R, Radhika K, Sivagnanam S, Rajeswari R, Hemalatha S, Kanimozhi A. An appraisal of available nutrients status and soil fertility mapping for Salem District of Tamil Nadu. *Madras Agricultural Journal*. 2014;101(1-3):51-58.
- Nayak DR, Babu S, Ramasamy P. Assessment of phosphorus availability in different soil types of Tamil Nadu. *Indian Journal of Agricultural Research*. 2019;53(6):705-713.
- Taghizadeh-Mehrjardi R, Minasny B. Integrating remote sensing data with Random Forest to map soil properties in arid regions. *Geoderma*. 2021;406:115553.

How to cite this article: Reddy BB, Maragatham S, Santhi R, Balachandrar D, Lakshmi DV, Davamani V, et al. Digital mapping of available phosphorus in agricultural soils of Tamil Nadu using Random Forest and environmental covariates. *International Journal of Advanced Biochemistry Research*. 2026;SP-10(10):522-527. doi: <https://www.doi.org/10.33545/26174693.2026.v10.i10Sg.10333>

Authors

B Bhanukiran Reddy

Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

S Maragatham

Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

R Santhi

Tamil Nadu Agricultural University, Coimbatore Tamil Nadu, India

D Balachandrar

Tamil Nadu Agricultural University, Coimbatore Tamil Nadu, India

D Vijaya Lakshmi

Tamil Nadu Agricultural University, Coimbatore Tamil Nadu, India

V Davamani

Tamil Nadu Agricultural University, Coimbatore Tamil Nadu, India

D Vasu

Tamil Nadu Agricultural University, Coimbatore Tamil Nadu, India

M Gopalakrishnan

Tamil Nadu Agricultural University, Coimbatore Tamil Nadu, India