



Geospatial Soil Texture Prediction in Cebu, Philippines: A Comparative Study of UK, MLR, and ANN

Jeremy P. Mondejar^{1*}, and Hershey P. Mondejar²

¹ Institute of Agricultural and Biosystems Engineering, Cebu Technological University – Barili Campus, Barili, Cebu, 6036, Philippines

² College of Veterinary Medicine, Cebu Technological University – Barili Campus, Barili, Cebu, 6036, Philippines

* Correspondence: jeremy.mondejar@ctu.edu.ph

Citation:

Mondejar, J.P.; Mondejar, H.P. Geospatial soil texture prediction in Cebu, Philippines: A comparative study of UK, MLR, and ANN. *ASEAN J. Sci. Tech. Report.* **2026**, *29*(8), e265019. <https://doi.org/10.55164/ajstr.v29i8.265019>

Article history:

Received: April 17, 2026

Revised: May 31, 2026

Accepted: June 21, 2026

Available online: July 5, 2026

Publisher's Note:

This article is published and distributed under the terms of the Thaksin University.

Abstract: Many soil maps in the Philippines are outdated, lacking the detail required for effective land management and agricultural planning. Traditional survey methods, still widely used, often fail to capture the spatial variability of soil properties, which is critical given the country's diverse topography and land-use patterns. This study compares three models—Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), and Universal Kriging (UK) for soil texture classification, using open-source software tools: R and SAGA GIS. Results indicate that the UK model significantly outperforms both MLR and ANN, with the highest overall accuracy of 77.25%, a strong kappa value (0.5757), and a statistically significant p-value ($5.937e-10$). The UK model excels in class-wise performance, showing superior sensitivity for key classes such as Class 1 (0.9576) and Class 6 (0.6719), while maintaining high specificity across all classes. Its balanced accuracy, as high as 0.8949 for Class 1, reflects its ability to handle both major and minority classes effectively. In contrast, MLR and ANN show lower sensitivity, often failing to classify minority classes (0 sensitivity in multiple cases) and achieving poorer balanced accuracy overall. The findings highlight the UK as the most reliable classifier among the three, excelling in both overall and class-specific performance. Future research should focus on further optimizing the UK model and testing its applicability to other datasets and classification tasks. Additionally, expanding the dataset with more samples from underrepresented classes could further improve model accuracy.

Keywords: Predictive DSM; soil texture classification; geostatistical modeling; machine learning

1. Introduction

The application of predictive digital soil mapping (PDSM) in the Philippine context presents several research gaps that need to be addressed to enhance the effectiveness and accuracy of soil mapping efforts. These gaps stem from the Philippines' unique geographical, climatic, and socio-economic conditions, which can significantly influence soil properties and their distribution. One of the primary gaps is the limited availability of high-resolution soil data across the archipelago. Many existing soil maps in the Philippines are outdated and lack the granularity required for effective land management and agricultural planning. For instance, while digital soil mapping techniques have been successfully applied in other regions, the Philippines still relies heavily on traditional soil survey methods that often fail to capture soil

property spatial variability adequately [1]. This is particularly concerning given the diverse topography and land use practices across the islands, which can lead to significant variations in soil characteristics.

Moreover, there is a need to integrate advanced modeling techniques, such as multiple linear regression (MLR), machine learning, and geostatistics, into PDSM practices in the Philippines. However, the application of these techniques in the Philippine context remains limited. The potential to use remote sensing data and environmental covariates to enhance predictions of soil properties has not been fully explored, which could significantly improve the accuracy of soil maps [2, 3]. Multiple linear regression is a well-established statistical technique widely used for predictive modeling in digital soil mapping [1, 4]. It models the relationship between a dependent variable (e.g., soil property) and multiple independent variables (e.g., environmental covariates). MLR's simplicity and interpretability make it suitable for initial exploratory analysis and establishing baseline predictive performance. Artificial neural networks (ANNs), a machine learning method, have been extensively applied to predictive digital soil mapping due to their ability to capture complex nonlinear relationships between soil properties and environmental covariates [5]. ANNs can learn from data without explicit programming, making them well-suited for modeling the inherent complexity of soil systems [6].

Universal kriging (UK) is a geostatistical technique that incorporates both spatial autocorrelation and environmental covariates to predict soil properties at unsampled locations [1]. The UK has been used in digital soil mapping to produce spatially continuous predictions while accounting for the spatial structure of soil properties [1]. The advantage of the UK is its ability to provide estimates of prediction uncertainty, which is crucial for decision-making in soil management. The findings will significantly enhance the accuracy and spatial resolution of soil information [4, 7, 8], providing a robust foundation for advanced crop and livestock land suitability modeling. By integrating predictive soil attributes with terrain and climate variables, the study will enable science-based identification of optimal areas for specific crops and agrosilvopastoral production systems [9–12]. This supports improved farm-level planning, efficient resource allocation, and risk reduction in agricultural investments [9–11]. Ultimately, the study promotes data-driven decision-making that advances sustainable agriculture, resilient food systems, and environmental conservation [10, 13] across Cebu Province.

This study aims to compare the predictive performance of MLR, ANN, and UK in PDSM by evaluating their ability to model soil texture using environmental covariates. This research seeks to determine the most effective method for accurate soil texture prediction by analyzing each model's strengths in handling linear relationships, capturing non-linear complexities, and incorporating spatial autocorrelation.

2. Materials and Methods

2.1 Materials

The Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) was accessed and processed using Google Earth Engine (GEE) for this study. The SRTM DEM, with a 30-meter spatial resolution, provides elevation data derived from radar interferometry, ensuring high accuracy and broad coverage for topographic analysis. The processed DEM was used as a reliable input for terrain analysis, enabling the extraction of key derivatives, including slope, aspect, and hydrological features (Table 1).

2.2 The study area

Cebu is the second-largest metropolitan area in the Philippines, located in the central Visayas, with a population exceeding 4 million [14]. It features diverse landscapes ranging from coastal plains to mountainous terrain [15] and has a tropical climate characterized by a dry season (March–May) and a wet season (June–November) [16, 17]. Agriculture remains a key economic activity, with major products including sugarcane, coconut, and food crops [18]. However, the region faces land use change, soil degradation, deforestation, and climate-related challenges that threaten environmental sustainability [19]. Cebu's complex topography is strongly shaped by karst formations, particularly in central areas [15], and further influenced by vegetation patterns and urbanization, especially in Cebu City [16]. Topographic variation also controls soil distribution and texture, which are critical for agriculture and land management [1]. Given its diverse geography, climate, and socio-economic conditions, Cebu provides a suitable setting for predictive digital soil mapping to support sustainable soil resource management and informed decision-making.

2.3 Study framework

This study followed a predictive digital soil mapping (DSM) framework integrating field soil observations, terrain-derived environmental covariates, statistical analysis, machine-learning techniques, and geostatistical modeling. The workflow consisted of four major phases: (i) data acquisition and preprocessing, (ii) exploratory data analysis and transformation, (iii) predictive modeling and spatial interpolation, and (iv) model validation and final map generation.

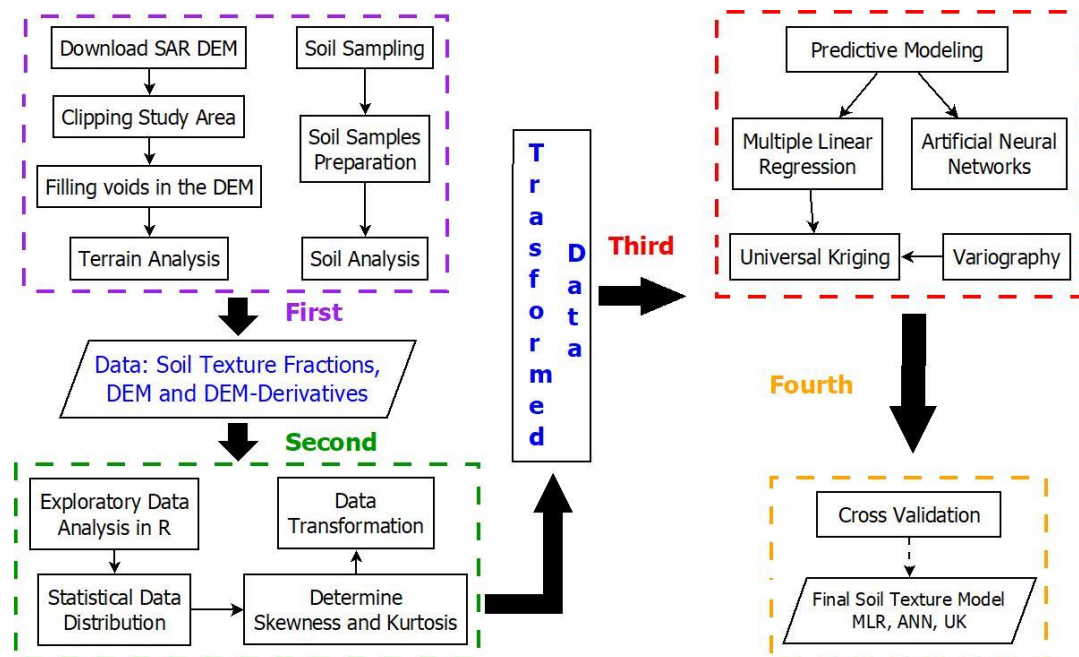


Figure 1. Workflow of predictive digital soil mapping showing data acquisition, exploratory analysis and transformation, predictive modeling (MLR, ANN, and universal kriging), and cross-validation leading to final soil texture maps.

2.4 Digital elevation model and terrain derivatives

A synthetic aperture radar (SAR)–derived digital elevation model (DEM) was acquired and preprocessed (clipped, void- and sink-filled) to ensure hydrological consistency (Table 1). Geomorphometric analysis generated terrain covariates relevant to soil formation, including slope, aspect, curvature, and related indices. Mountain ridge proximity was additionally examined following [1]. These DEM-derived variables were used as continuous environmental predictors and are widely recognized as key controls on soil texture distribution [1, 5, 20–22].

2.5 Soil sampling and laboratory analysis

Soil samples were collected from representative locations using a combined convenience (legacy-based) and stratified sampling approach to adequately capture landscape variability. A total of 445 samples were obtained across the City of Carcar (22 samples) and the municipalities of Barili (371 samples) and Sibonga (52 samples) (Figure 2). Collected samples were air-dried, processed, and sieved before laboratory analysis. Soil texture fractions—sand, silt, and clay—were determined using standard analytical procedures at the Department of Agriculture–Region VII Soils Laboratory.

2.6 Exploratory data analysis and transformation

Exploratory data analysis using R, including summary statistics, graphical diagnostics, and distribution plots, was performed to identify outliers and assess variability of soil texture fractions. Skewness and kurtosis were computed to evaluate departures from normality, and, where needed, Box–Cox–type power transformations were applied to reduce skewness, stabilize variance, and better satisfy linear regression and geostatistical assumptions [1, 23, 24]. Because the response variables were transformed to meet the normality

assumption, all identified best model outputs were subsequently back-transformed to their original scale for interpretation. Furthermore, predictor variables were normalized via min-max scaling to reduce the impact of different ranges of predictor values [23, 25].

Table 1. Terrain derivatives generated from the digital elevation model (DEM) and organized according to the SCORPAN environmental covariate framework to represent relief (R) and spatial position (N) factors in predictive digital soil mapping.

Attribute type	DEM derivatives
Basic terrain morphology	Elevation, Slope, Gradient, Slope Height, Slope Length, Standardized Height, Normalized Height
Terrain shape and curvature	Plan Curvature, Profile Curvature, Convergence Index, Gradient Difference
Hydrological terrain attributes	Flow Accumulation, Total Catchment Area, Flow Width, Maximum Flow Path Length, Topographic Wetness Index, LS-Factor, Channel Network Distance, Distance Channels Proximity, Stream Proximity
Terrain position indices	Mountain Ridge Proximity, Topographic Position Index, Mid-Slope Position, Valley Depth, Landform
Multiscale relief structure indices	MRVBF, MRRTF, Melton Ruggedness Number
Energy and terrain exposure proxies	Analytical Hillshading, Diurnal Anisotropic Heating
Composite/secondary terrain indices	Entropy, Confusion Index, Maximum Membership, Mass Balance Index
Spatial position factors	Latitude, Longitude

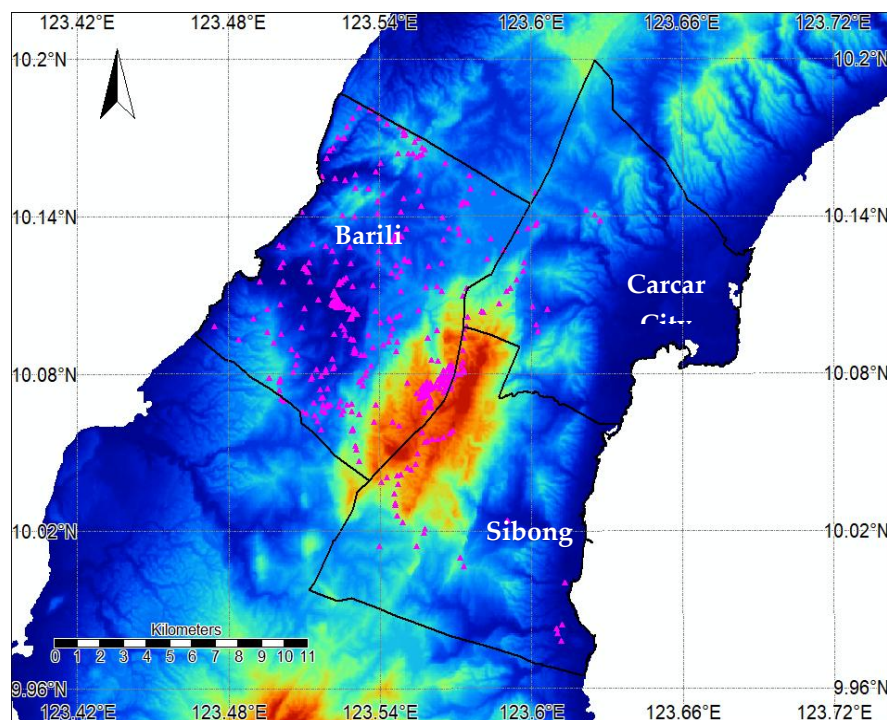


Figure 2. Spatial distribution of the 445 soil sampling locations across the City of Carcar and the municipalities of Barili and Sibonga, Cebu, collected using a combined convenience and stratified sampling approach.

2.7 Integration of soil and environmental data

Soil texture observations were spatially linked with corresponding terrain-derived covariates extracted from the DEM. The resulting georeferenced soil texture data were used as point-based response variables for predictive modeling [1, 2, 26, 27]. This integration resulted in a modeling dataset in which each soil sample was associated with a set of environmental predictors representing landscape controls on soil variability.

2.8 Predictive modeling approaches

Soil texture fractions were predicted using multiple modeling techniques. Multiple linear regression (MLR) provided an interpretable baseline linking terrain-derived covariates to soil texture [1, 28, 29]. Artificial neural networks (ANN) were applied to capture nonlinear relationships between environmental predictors and soil texture fractions [29–31]. Geostatistical modeling using variogram analysis (Table 2) and universal kriging (UK) incorporated spatial autocorrelation, combining environmental trends with spatial dependence to identify spatially explicit predictors from MLR as external drift parameters for spatial interpolation [1, 31].

Table 2. The variogram parameters of soil texture fractions are applied to the exponential model.

Soil Texture fraction	Semivariogram model	Nugget effect	Sill	Nugget-to-sill ratio	Effective range
Clay	exponential	400	2158	0.185	3450
Sand	exponential	0.4	2.57	0.156	5215
Silt	exponential	0.04	2.24	0.018	1798

2.9 Model comparison based on confusion-matrix-derived parameters

Model performance was evaluated using 5-fold cross-validation, in which 80% of the data were used for training and the remaining 20% for validation in each fold. Standard random k-fold CV may overestimate accuracy when training and test observations are spatially proximate; spatial cross-validation is recommended in future work to provide conservative, unbiased estimates for a geostatistical model. Comparative evaluation of the classification models was conducted using parameters derived from the confusion matrix to capture both overall and class-specific performance. For each model, a confusion matrix was generated on the independent validation data, summarizing true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each soil texture class [32–34]. From these counts, standard discrimination metrics were computed, including sensitivity/recall ($TP / (TP + FN)$), specificity ($TN / (TN + FP)$), precision or positive predictive value ($TP / (TP + FP)$), negative predictive value ($TN / (TN + FN)$), and overall accuracy ($(TP + TN) / (TP + FP + TN + FN)$) [33, 35, 36]. Because the dataset exhibited substantial class imbalance, additional emphasis was placed on balanced accuracy (the mean of sensitivity and specificity) and class-averaged scores, which are less biased by the majority classes than overall accuracy [33, 34, 36]. Models were compared by examining: (i) overall accuracy as a broad indicator of performance; (ii) per-class sensitivity and specificity to identify classes that were consistently misclassified or confused; and (iii) balanced accuracy to assess performance under class imbalance and to avoid over-optimistic assessments based solely on accuracy [33, 36]. This confusion-matrix-based comparison enabled a nuanced evaluation of each model's strengths and weaknesses across soil texture classes and provided the basis for selecting the preferred model for subsequent spatial prediction and mapping.

2.10 Synthesis of Sand, Silt, and Clay Predictions into Final Soil Texture Model

Soil texture was classified into 12 USDA classes: Class 1 = Clay, Class 2 = Silty Clay, Class 3 = Silty Clay Loam, Class 4 = Sandy Clay, Class 5 = Sandy Clay Loam, Class 6 = Clay Loam, Class 7 = Silt, Class 8 = Silt Loam, Class 9 = Loam, Class 10 = Sand, Class 11 = Loamy Sand, Class 12 = Sandy Loam. The final soil texture model was generated in SAGA GIS using the Soil Texture Classification module, integrating the back-transformed best-performing grid models of sand, silt, and clay. This process produced continuous spatial predictions of soil texture classes across the study area, providing spatially explicit outputs suitable for land evaluation, agricultural planning, and environmental modeling.

3. Results

3.1 Performance comparison of MLR, ANN, & UK

The study compares the performance of three methods (MLR, ANN, and UK) to evaluate their effectiveness in modeling the spatial distribution of topsoil texture. Model performance was evaluated using accuracy, sensitivity, specificity, kappa, and balanced accuracy to assess predictive capability across both dominant and minority classes.

3.1.1 Sensitivity (Recall)

Sensitivity results indicate that all models most effectively identified the clay class, as shown in Table 3. The MLR model achieved high sensitivity for Class 1 (clay; 0.8410) but failed to correctly identify most other classes, including loam (Class 9), for which sensitivity was 0.0000. The ANN model showed moderate sensitivity to clay (0.6290) but was insensitive to several classes, including silty clay (Class 2) and loam (Class 9). The UK model achieved the highest sensitivity for clay (0.9576) and also showed good sensitivity for clay loam (Class 6; 0.6719), while sand (Class 10) and loamy sand (Class 11) were not correctly identified.

Table 3. Sensitivity (recall) of MLR, ANN, and UK models for soil texture classes, highlighting the best- and worst-performing classes based on the ability to identify true class occurrences correctly.

Model	Best Performing Classes	Worst Performing Classes
MLR	Class 1 _{clay} (0.8410)	Most others (0.0000, e.g., Class 9 _{loam})
ANN	Class 1 _{clay} (0.6290)	Several classes (0.0000, e.g., 2 _{silty clay} , 9 _{loam})
UK	Class 1 _{clay} (0.9576), Class 6 _{clay loam} (0.6719)	Classes 10 _{sand} , 11 _{loamy sand} (0.0000)

3.1.2 Specificity

Specificity was generally high across all models, as shown in Table 4. The MLR model showed perfect specificity (1.0000) for several less frequent classes, including silty clay, sandy clay, sand, and loamy sand. The ANN model demonstrated good overall specificity, with particularly high performance for clay loam (0.9658). The UK model exhibited consistently high specificity values, close to 1.0000, across all soil texture classes evaluated.

Table 4. Specificity of MLR, ANN, and UK models across soil texture classes, indicating the ability of each model to identify non-class instances correctly.

Model	Specificity Performance
MLR	High for rare classes (e.g., 1.0000 in Classes 2 _{silty clay} , 4 _{sandy clay} , 10 _{sand} , 11 _{loamy sand})
ANN	Good overall, notably Class 6 _{clay loam} (0.9658)
UK	Consistently high (close to 1) for all classes (1 _{clay} , 2 _{silty clay} , 4 _{sandy clay} , 5 _{sandy clay loam} , 6 _{clay loam} , 9 _{loam} , 10 _{sand} , 11 _{loamy sand} , 12 _{sandy loam})

3.1.3 Balanced accuracy

Balanced accuracy revealed clearer differences among models by accounting for both sensitivity and specificity, as shown in Table 5. The MLR model achieved its highest balanced accuracy in clay (0.6565), while silty clay, loam, and loamy sand showed balanced accuracies close to 0.5000, indicating chance-level performance. The ANN model performed best for sandy clay loam (0.5293), with several classes, including silty clay, loam, and loamy sand, again showing balanced accuracy near 0.5000. The UK model achieved the highest balanced accuracy overall, particularly for clay (0.8949) and clay loam (0.7859), whereas sand and loamy sand remained at chance-level performance.

3.1.4 Overall model performance

Table 6 shows that the overall accuracy was highest for the UK model (0.7725), with a 95% confidence interval of 0.7307–0.8107, and exceeded the no information rate (0.6374), as reflected by a statistically significant p-value (5.94E-10). The MLR model achieved moderate accuracy (0.5856), while the ANN model had the lowest accuracy (0.4347); neither model outperformed the no-information rate. Agreement, measured

by kappa, followed the same pattern, with UK showing moderate agreement (0.5757), MLR showing low agreement (0.1703), and ANN showing minimal agreement (0.071).

Table 5. Balanced accuracy of MLR, ANN, and UK models by soil texture class, combining sensitivity and specificity to account for class imbalance.

Model	Best Classes	Worst Classes
MLR	Class 1 _{clay} (0.6565)	Class 2 _{silty clay} , 9 _{loam} , 11 _{loamy sand} (~0.5000)
ANN	Class 5 _{sandy clay loam} (0.5293)	Class 2 _{silty clay} , 9 _{loam} , 11 _{loamy sand} (~0.5000)
UK	Class 1 _{clay} (0.8949), Class 6 _{clay loam} (0.7859)	Classes 10 _{sand} , 11 _{loamy sand} (0.5000)

Table 6. Overall classification performance of MLR, ANN, and UK models based on accuracy, confidence intervals, no information rate, statistical significance, and kappa agreement.

Metric	MLR	ANN	UK
Accuracy	0.5856	0.4347	0.7725
95% Confidence Interval	(0.5382, 0.6318)	(0.3880, 0.4822)	(0.7307, 0.8107)
No Information Rate	0.6374	0.6374	0.6374
P-Value (Acc > NIR)	0.9893	1	5.94E-10
Kappa	0.1703	0.071	0.5757

3.2 Soil texture model by universal kriging

The spatial distribution of topsoil types across municipalities in Cebu province is predicted to show greater diversity from south to north than from east to west (Figure 3). Dumanjug (72%), Sibonga (80%), Barili (82%), Carcar City (82%), Ronda (83%), Pinamungajan (84%), Alcantara (87%), Aloguinsan (93%), Samboan (96%), and Santander (99%) have predominantly clay soil due to its high percentage in these municipalities. Clay Loam-dominated municipalities are San Remegio (76%) and Pilar (86%). Municipalities with a mixed dominance of two soil types summing to at least 80% include Bogo City (46% clay, 34% clay loam), Borbon (42% clay, 44% clay loam), Santa Fe (51% clay, 29% sandy loam), Medellin (26% loam, 62% sandy loam), Mandaue City (21% loamy sand, 61% sandy loam), Tabogon (30% clay, 54% clay loam), San Francisco (32% clay, 60% clay loam), Sogod (36% clay, 60% clay loam), Minglanilla (40% clay, 54% clay loam), Tabuelan (58% clay loam, 31% loam), Malabuyoc (59% clay, 21% clay loam), Naga (59% clay, 35% clay loam), San Fernando (57% clay, 40% clay loam), Ginatilan (63% clay, 18% sandy clay), Tudela (63% sandy clay loam, 30% clay loam), and Moalboal (69% clay, 22% clay loam). The majority of these mixed soil types are clay and clay loam.

Municipalities with significant and unique soil type dominance as loam of Tuburan (42%), as silty clay-dominated of Madrideojos (95%) and Bantayan (81%), and as sandy clay loam of Poro (47%), Alegria (64%), Tudela (63%), Cebu City (38%), Compostela (42%), Consolacion (46%), Dalaguete (35%), and Liloan (38%). Others, such as Asturias (19% loamy sand, 64% sandy loam), Balamban (56% sandy loam), Carmen (48% sandy loam), Cordova (85% loamy sand), Daanbantayan (39% loamy sand, 57%, sandy loam), Danao City (50% sandy loam), Liloan (40% sandy loam), Lapu-Lapu City (52% loamy sand), Mandaue City (21% loamy sand, 61% sandy loam), and Medellin (62% sandy loam) show a significant percentage of sandier textures. The model's performance, although with 77% accuracy, suggests it has limitations in accurately predicting the spatial distribution of topsoil texture across the Province of Cebu, particularly for underrepresented classes (Figure 3). Hypothetically, Cebu's topsoil texture exhibits a diverse spatial pattern, with sandy loam (SL) prevalent from highland to lowland and down to the coastal areas in Central Cebu (Cebu City, Mandaue City, Consolacion, Liloan, Compostela, Danao City, Carmen, Catmon, Asturias, and Balamban), Daanbantayan, and Medellin. Clay soil dominates upland and mountainous regions in the northern and southern parts. Sandy clay loam is estimated to be dominant in the mountainous areas of Argao, Badian, Dalaguete, Alegria, and Alcoy. Loam soil is concentrated in the lowlands of northern Cebu (Tabuelan, Tuburan, Carmen, Catmon, and Medellin). Loamy sand dominates greatly in Cordova, Lapu-Lapu City, and Daanbantayan. Olango Island in Lapu-Lapu City is expected to have entirely sandy soil. Despite its limitations, the model could still be a valuable tool

for preliminary soil mapping and agricultural planning, offering insights into erosion risk, water retention, and crop suitability.

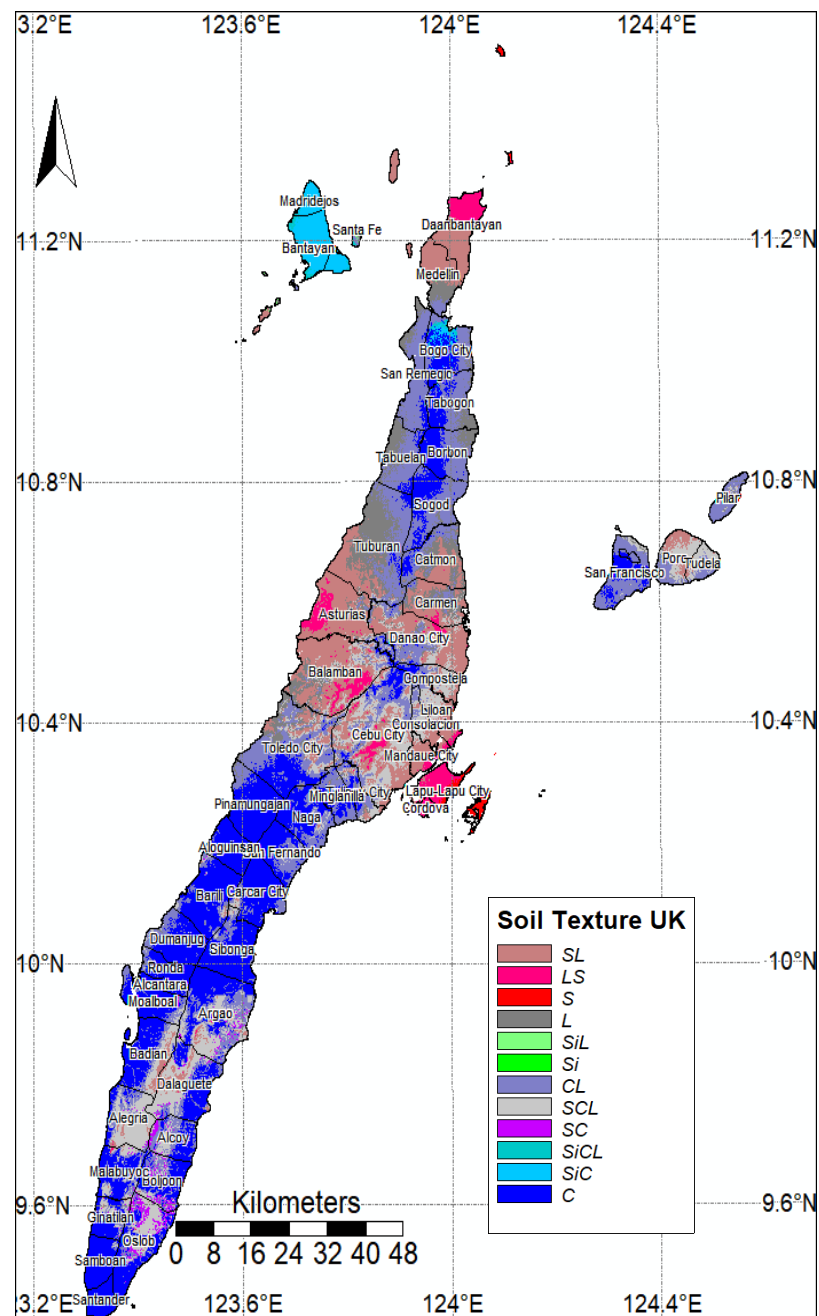


Figure 1. Universal kriging model of the spatial distribution of topsoil texture in the Province of Cebu by predictive digital soil mapping.

4. Discussion

The comparative evaluation of MLR, ANN, and UK highlights how both model type and class imbalance shape soil texture classification performance. Similar to other digital soil mapping studies, performance was strongly class dependent, with higher recall for clayey classes and near-chance performance for several minority or spectrally ambiguous classes [5, 6, 36]. The consistently high sensitivity of clay across all models aligns with reports that clay-rich soils often exhibit clearer spectral and topographic signatures than soils with intermediate textures, facilitating discrimination in both remote-sensing-based and terrain-based

models [24, 27, 36]. Conversely, sandy and loamy classes, particularly loam and loamy sand in this study, showed very low sensitivity and balanced accuracy, reflecting well-documented challenges in separating intermediate or mixed textures, where transitions are gradual, and feature spaces overlap [5, 6, 27, 36]. High specificities for several rare classes, especially under MLR, likely result from strong class imbalance: models rarely predict these classes, yielding few false positives but many false negatives. This pattern is consistent with studies demonstrating that overall accuracy and specificity can overestimate performance on imbalanced soil texture datasets, and that metrics incorporating both omission and commission errors (e.g., kappa, balanced accuracy) provide a more realistic assessment [5, 24, 36]. The use of balanced accuracy in this study, therefore, follows current recommendations for evaluating imbalanced soil texture classification problems [24].

The relatively poor performance of ANN in this study is consistent with the findings of Mallah et al. (2023), who reported an accuracy of 52% and a kappa coefficient of 0.38. This result is also consistent with broader evidence suggesting that ANNs tend to underperform when applied to small-to-medium-sized datasets, where overfitting and sensitivity to hyperparameters reduce generalization [5, 23, 30]. The superior overall accuracy and kappa of the UK relative to MLR and ANN suggest that explicitly modeling spatial structure can be advantageous when covariates or sample size are insufficient for complex machine learning models. Comparable results have been reported, showing that universal kriging outperformed MLR for texture fractions, especially with limited sample sizes and strong spatial autocorrelation [1]. The superior performance of the UK can be attributed to its ability to account for local nonstationarity [1] and incorporate the spatial structure [37] of soil properties, which are crucial for accurate prediction. Furthermore, the UK model used in this study was a multivariate model that included the significant predictors identified by MLR. A third key limitation is the severe geographic imbalance (83% of samples from Barili). Cross-validation accuracy (UK: 77.25%) was estimated within the sampled domain; in unsampled municipalities, UK effectively reduces to MLR behavior, eliminating its spatial autocorrelation advantage. Spatially-blocked cross-validation is recommended for future province-wide DSM studies to provide realistic accuracy estimates [3, 12]. The weak performance for certain classes across all models underscores two key limitations: (i) skewed class distributions, which have been repeatedly shown to depress recall for rare texture classes [5, 6, 24], and (ii) the inherent difficulty of discretizing a continuous particle-size space into categorical texture classes [24, 38]. Approaches such as indirect classification via predicted particle-size fractions [24, 26, 27], and targeted sampling to bolster rare classes [5] have all been proposed to mitigate these issues and could further improve performance in future work.

5. Conclusions

This study compared multiple linear regression (MLR), artificial neural networks (ANN), and universal kriging (UK) for predicting soil texture classes using terrain-derived covariates. The UK attained the highest overall accuracy (0.7725) and kappa (0.5757), significantly outperforming the no-information rate, whereas MLR (0.5856) and ANN (0.4347) showed weak agreement and did not exceed baseline performance. All models predicted clay best, with UK demonstrating particularly strong sensitivity and balanced accuracy for clay and good performance for clay loam. However, sand and loamy sand showed near-chance-level performance. High specificity but low sensitivity for minority classes indicates limited detection of underrepresented textures. Balanced accuracy proved more informative than overall accuracy under class imbalance, reinforcing the UK's superiority and the near-random behavior of MLR and ANN for several classes. Based on these findings, the UK is recommended for operational soil texture mapping in similar settings due to its ability to model spatial autocorrelation. Future sampling should target underrepresented classes through stratified or focused designs to reduce imbalance. Model evaluation should prioritize class-wise metrics (sensitivity, specificity, balanced accuracy, and kappa) rather than relying solely on overall accuracy. Further research should explore hybrid spatial-machine learning approaches, imbalance-aware strategies, and expanded covariate sets to improve discrimination of minority classes. The results highlight the importance of modeling spatial dependence in digital soil mapping. Reliable texture maps can enhance decision-making in precision agriculture, irrigation, erosion control, and land-use planning, including targeted soil amendment strategies for problematic textures such as sandy or saline soils. However,

uncertainty in minority classes underscores the need for additional field validation or denser sampling in affected areas. Overall, improvements in predictive mapping depend not only on algorithm choice but also on sampling design, data quality, and appropriate evaluation metrics.

5. Acknowledgments

The authors gratefully acknowledge the Soils Laboratory of the Department of Agriculture – Regional Field Office VII for conducting the laboratory analyses of the soil samples used in this study. The authors also extend their sincere appreciation to Cebu Technological University (CTU)–Barili Campus for providing institutional and research support that facilitated the implementation of this work.

Author Contributions: JM: Conceptualization; Methodology; Software; Formal analysis; Data curation; Visualization; Writing – original draft; Writing – review & editing; Supervision; Project administration.

HM: Investigation; Resources; Validation; Data curation; Writing – original draft; Writing – review & editing.

Funding: This study was funded by Cebu Technological University through its General Appropriations Act (GAA) research fund, and the university also funded the APC.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results”.

References

- [1] Mondejar, J. P.; Tongco, A. F. Estimating Topsoil Texture Fractions by Digital Soil Mapping - a Response to the Long Outdated Soil Map in the Philippines. *Sustain. Environ. Res.* **2019**, 29(1), 31. <https://doi.org/10.1186/s42834-019-0032-5>
- [2] Azizi, K.; Garosi, Y.; Ayoubi, S.; Tajik, S. Integration of Sentinel-1/2 and Topographic Attributes to Predict the Spatial Distribution of Soil Texture Fractions in Some Agricultural Soils of Western Iran. *Soil Tillage Res.* **2023**, 229, 105681. <https://doi.org/10.1016/j.still.2023.105681>
- [3] Khammao, P.; Rattanapichai, W.; Pitakdantham, R.; Kasemsap, P.; Sajjaphan, K. The Potential of Near-infrared Spectroscopy to Predict Soil Nutrient Contents Based on Soil Color. *ASEAN J. Sci. Technol. Rep.* **2024**, 27(5), e252637. <https://doi.org/10.55164/ajstr.v27i5.252637>
- [4] Forkuor, G.; Hounkpatin, O. K. L.; Welp, G.; Thiel, M. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLOS ONE* **2017**, 12(1), e0170478. <https://doi.org/10.1371/journal.pone.0170478>
- [5] Mallah, S.; Delsouz Khaki, B.; Davatgar, N.; Poppiel, R. R.; Demattê, J. A. M. Digital Mapping of Topsoil Texture Classes Using a Hybridized Classical Statistics–Artificial Neural Networks Approach and Relief Data. *AgriEngineering* **2023**, 5(1), 40–64. <https://doi.org/10.3390/agriengineering5010004>
- [6] Sabbaghi, M. A.; Esfandiari, M.; Eftekhari, K.; Torkashvand, A. M. Predictive Map of Soil Texture Classes Using Decision Tree Model and Neural Network with Features of Geomorphology Level. *Can. J. Soil Sci.* **2024**, 104(1), 72–90. <https://doi.org/10.1139/cjss-2023-0011>
- [7] Liu, J.; Yang, K.; Tariq, A.; Lu, L.; Soufan, W.; El Sabagh, A. Interaction of Climate, Topography and Soil Properties with Cropland and Cropping Pattern Using Remote Sensing Data and Machine Learning Methods. *Egypt. J. Remote Sens. Space Sci.* **2023**, 26(3), 415–426. <https://doi.org/10.1016/j.ejrs.2023.05.005>
- [8] Wang, Y.; Yuan, F.; Cammarano, D.; Liu, X.; Tian, Y.; Zhu, Y.; Cao, W.; Cao, Q. Integrating Machine Learning with Spatial Analysis for Enhanced Soil Interpolation: Balancing Accuracy and Visualization. *Smart Agric. Technol.* **2025**, 11, 101032. <https://doi.org/10.1016/j.atech.2025.101032>
- [9] Dornik, A.; Chețan, M. A.; Crișan, T. E.; Heciko, R.; Gora, A.; Drăguț, L.; Panagos, P. Geospatial Evaluation of the Agricultural Suitability and Land Use Compatibility in Europe’s Temperate Continental Climate Region. *Int. Soil Water Conserv. Res.* **2024**, 12(4), 908–919. <https://doi.org/10.1016/j.iswcr.2024.01.002>

- [10] Sperandio, H. V.; de Morais, M. S.; de Jesus França, L. C.; Mucida, D. P.; Santana, R. C.; da Silva, R. S.; Rodrigues, C. R.; de Faria, B. L.; de Azevedo, M. L.; Gorgens, E. B. Land Suitability Modeling Integrating Geospatial Data and Artificial Intelligence. *Agric. Syst.* **2025**, *223*, 104197. <https://doi.org/10.1016/j.agry.2024.104197>
- [11] Taghizadeh-Mehrjardi, R.; Nabiollahi, K.; Rasoli, L.; Kerry, R.; Scholten, T. Land Suitability Assessment and Agricultural Production Sustainability Using Machine Learning Models. *Agronomy* **2020**, *10*(4), 573. <https://doi.org/10.3390/agronomy10040573>
- [12] Tongkaemkaew, U.; Commons, M.; Verhofste, R.; Phooi, C. L.; Wonglom, P.; Bua-Khwan, B. Regeneration of Soil Fertility in Relationship with the Diversification of Rubber Agroforestry Systems. *ASEAN J. Sci. Technol. Rep.* **2025**, *28*(2), e255018. <https://doi.org/10.55164/ajstr.v28i2.255018>.
- [13] Bhullar, A.; Nadeem, K.; Ali, R. A. Simultaneous Multi-Crop Land Suitability Prediction from Remote Sensing Data Using Semi-Supervised Learning. *Sci. Rep.* **2023**, *13*(1), 6823. <https://doi.org/10.1038/s41598-023-33840-6>
- [14] Philippine Statistics Authority (PSA). 2020 Census of Population and Housing: Cebu Province; PSA: Manila, Philippines, 2022. Available online: <https://psa.gov.ph> (accessed May 2026).
- [15] Lumongsod, R. M. G.; Ramos, N. T.; Dimalanta, C. B. Mapping the Karstification Potential of Central Cebu, Philippines Using GIS. *Environ. Earth Sci.* **2022**, *81*(18), 449. <https://doi.org/10.1007/s12665-022-10572-w>
- [16] Cañete, S. F.; Schaap, L. L.; Andales, R.; Otadoy, R. E. S.; Blanco, A. C.; Babaan, J.; Cruz, C. Analysis of the Impact of Vegetation Distribution, Urbanization, and Solar Radiation on the Seasonal Variation of the Urban Heat Island Effect in Cebu City Using Landsat and Global Horizontal Irradiance Data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-4-W19*, 93–100. <https://doi.org/10.5194/isprs-archives-XLII-4-W19-93-2019>
- [17] Edillo, F.; Ymbong, R. R.; Navarro, A. O.; Cabahug, M. M.; Saavedra, K. Detecting the Impacts of Humidity, Rainfall, Temperature, and Season on Chikungunya, Dengue and Zika Viruses in Aedes Albopictus Mosquitoes from Selected Sites in Cebu City, Philippines. *Virol. J.* **2024**, *21*(1), 42. <https://doi.org/10.1186/s12985-024-02310-4>
- [18] Galang, W. N.; Tabañag, I. D.; Loretero, M. RS-GIS Approach on Biomass Energy Potential Estimation of Sugarcane Residues in Medellin, Cebu. *Appl. Environ. Res.* **2022**, *44*(1). <https://doi.org/10.35762/AER.2021.44.1.6>
- [19] Lillo, E. P.; Malaki, A. B.; Alcazar, S. M. T.; Rosales, R.; Redoblado, B. R.; Diaz, J. L. B.; Pantinople, E. M.; Buot, I. E. J. Inventory of Native and Mother Trees in Key Biodiversity Areas of Cebu Island, Philippines for Species Selection in Local Reforestation Programs. *Biodiversitas J. Biol. Divers.* **2021**, *22*(11). <https://doi.org/10.13057/biodiv/d221105>
- [20] Kasraei, B.; Schmidt, M. G.; Zhang, J.; Bulmer, C. E.; Filatow, D. S.; Arbor, A.; Pennell, T.; Heung, B. A Framework for Optimizing Environmental Covariates to Support Model Interpretability in Digital Soil Mapping. *Geoderma* **2024**, *445*, 116873. <https://doi.org/10.1016/j.geoderma.2024.116873>
- [21] Liu, F.; Zhang, G.-L.; Song, X.; Li, D.; Zhao, Y.; Yang, J.; Wu, H.; Yang, F. High-Resolution and Three-Dimensional Mapping of Soil Texture of China. *Geoderma* **2020**, *361*, 114061. <https://doi.org/10.1016/j.geoderma.2019.114061>
- [22] Mousavi, A.; Karimi, A.; Alavipanah, S. K.; Shahabi, M.; Safari, T. Spatial Modelling Soil Texture Variability in Northeastern Iran: Integrating Remote Sensing and Machine Learning. *Earth Surf. Process. Landf.* **2025**, *50*(8), e70110. <https://doi.org/10.1002/esp.70110>
- [23] Khaledian, Y.; Miller, B. A. Selecting Appropriate Machine Learning Methods for Digital Soil Mapping. *Appl. Math. Model.* **2020**, *81*, 401–418. <https://doi.org/10.1016/j.apm.2019.12.016>
- [24] Zhang, M.; Shi, W.; Xu, Z. Systematic Comparison of Five Machine-Learning Models in Classification and Interpolation of Soil Particle Size Fractions Using Different Transformed Data. *Hydrol. Earth Syst. Sci.* **2020**, *24*(5), 2505–2526. <https://doi.org/10.5194/hess-24-2505-2020>
- [25] Ali, P. J. M. Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor with Different Similarity Measurements. *ARO- Sci. J. Koya Univ.* **2022**, *10*(1), 85–91. <https://doi.org/10.14500/aro.10955>

- [26] Liu, J.; Ye, Y.; Wang, C.; Chen, S.; Jiang, Y.; Guo, X.; Jiang, Y. Machine Learning-Based Comparative Analysis on Direct and Indirect Mapping of Soil Texture Types Through Soil Particle Size Fractions Using Multi-Source Remote Sensing. *Agriculture* **2025**, *15*(13), 1395. <https://doi.org/10.3390/agriculture15131395>
- [27] Mirzaeitalarposhti, R.; Shafizadeh-Moghadam, H.; Taghizadeh-Mehrjardi, R.; Demyan, M. S. Digital Soil Texture Mapping and Spatial Transferability of Machine Learning Models Using Sentinel-1, Sentinel-2, and Terrain-Derived Covariates. *Remote Sens.* **2022**, *14*(23), 5909. <https://doi.org/10.3390/rs14235909>
- [28] Chagas, C. da S.; de Carvalho Junior, W.; Bhering, S. B.; Calderano Filho, B. Spatial Prediction of Soil Surface Texture in a Semiarid Region Using Random Forest and Multiple Linear Regressions. *CATENA* **2016**, *139*, 232–240. <https://doi.org/10.1016/j.catena.2016.01.001>
- [29] Taghizadeh-mehrjardi, R.; Toomanian, N.; Khavaninzadeh, A. R.; Jafari, A.; Triantafilis, J. Predicting and Mapping of Soil Particle-Size Fractions with Adaptive Neuro-Fuzzy Inference and Ant Colony Optimization in Central Iran. *Eur. J. Soil Sci.* **2016**, *67*(6), 707–725. <https://doi.org/10.1111/ejss.12382>
- [30] Feng, L.; Khalil, U.; Aslam, B.; Ghaffar, B.; Tariq, A.; Jamil, A.; Farhan, M.; Aslam, M.; Soufan, W. Evaluation of Soil Texture Classification from Orthodox Interpolation and Machine Learning Techniques. *Environ. Res.* **2024**, *246*, 118075. <https://doi.org/10.1016/j.envres.2023.118075>
- [31] Khanbabakhani, E.; Torkashvand, A. M.; Mahmoodi, M. A. The Possibility of Preparing Soil Texture Class Map by Artificial Neural Networks, Inverse Distance Weighting, and Geostatistical Methods in Gavoshan Dam Basin, Kurdistan Province, Iran. *Arab. J. Geosci.* **2020**, *13*(5), 237. <https://doi.org/10.1007/s12517-020-5134-1>
- [32] Farhadpour, S.; Warner, T. A.; Maxwell, A. E. Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices. *Remote Sens.* **2024**, *16*(3), 533. <https://doi.org/10.3390/rs16030533>
- [33] Luque, A.; Carrasco, A.; Martín, A.; de las Heras, A. The Impact of Class Imbalance in Classification Performance Metrics Based on the Binary Confusion Matrix. *Pattern Recognit.* **2019**, *91*, 216–231. <https://doi.org/10.1016/j.patcog.2019.02.023>
- [34] Tharwat, A. Classification Assessment Methods. *Appl. Comput. Inform.* **2020**, *17*(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- [35] Cabot, J. H.; Ross, E. G. Evaluating Prediction Model Performance. *Surgery* **2023**, *174*(3), 723–726. <https://doi.org/10.1016/j.surg.2023.05.023>
- [36] Obi, J. C. A Comparative Study of Several Classification Metrics and Their Performances on Data. *World J. Adv. Eng. Technol. Sci.* **2023**, *8*(1), 308–314. <https://doi.org/10.30574/wjaets.2023.8.1.0054>
- [37] Sabareeshwari, V.; Mary, P. C. N.; Mahendran, P. P.; Pandian, P. S.; Gurusamy, A.; Subhashini, R. Kriging: An Advanced Geostatistical Tool to Interpolate the Textural Variation Influences the Yield and Productivity of Tapioca (*Manihot Esculenta*) in Namakkal District, Tamil Nadu, India. *Int. J. Plant Soil Sci.*, **2021**, *33*(2), 53–59. <https://doi.org/10.9734/ijpss/2021/v33i230414>
- [38] Diana, N. A.; Widodo, T.; Saputri, N. D. Stabilization of Sandy Soil with High Salinity Conditions using Rice Husk Ash and Gypsum to Improve Physical and Mechanical Properties. *ASEAN J. Sci. Technol. Rep.* **2024**, *27*(4), e252961. <https://doi.org/10.55164/ajstr.v27i4.252961>