Soil Salinity Spatial Analysis to Develop a Machine Learning-Based Soil Resistivity Predictive Model

H.P.C.S. Manchanayake[1]*, H.M.B.S. Herath[2], Indrajanaka Makalanda[3]

[1] Post Graduate Student, Department of Geography, Faculty of Graduate Studies, University of Sri Jayewardenepura, Sri Lanka

[2] Senior Lecturer, Department of Geography, Faculty of Humanities and Social Sciences, University of Sri Jayewardenepura, Sri Lanka

[3] Senior Lecturer, Department of Decision Sciences, Faculty of Business, University of Moratuwa, Sri Lanka

chanaka.manchanayake@yahoo.com

ABSTRACT

The development of a sustainable and resilient electrical distribution network in a region optimizes the placement of electrical equipment and substations which reduces the risk of equipment failures and downtime and enhances cost effectiveness. However, the application of geospatial techniques with AI-based methods for these purposes is very limited in Sri Lanka. This study aims to develop a machine learning model-based predictive model by integrating geospatial techniques to identify the spatial distribution of soil salinity to detect soil resistivity to optimize suitable locations for earthing to enhance the efficiency of the electrical network. The study was carried out in the Rathmalana Electrical Engineer area, using images collected by the Sentinel-02 satellite acquired in the year 2022 and field-collected earth resistance values. Supervised learning algorithms such as linear regression and random forest were used to establish forecasting models. Their performances were measured using RMSE and R-squared. The obtained values of earth resistance varied, ranging from 0.02 to 52.00 ohms with a standard deviation of 6.40. The linear regression model achieved the best performance, followed by Lasso regression and the random forest algorithms, with lower errors with moderate accuracy. The study points out the need to achieve higher accuracy and proposes further exploration with alternative models by incorporation of more variables to achieve higher accuracy. The proposed model highlights the importance of understanding the resistivity and salinity of soil for designing effective earthing systems in other regions too.

Keywords: Soil salinity, Soil resistivity, Machine learning, Remote sensing, Predictive model

Introduction

Soil salinity is an important factor for many applications, such as agriculture, construction, and electrical engineering, among others. Proper knowledge regarding its impacts and proper methods for its detection and management are necessary to optimize the use of land and infrastructure development. Soil salinity refers to the concentration of soluble salts in the soil; the level of this salinity dramatically influences the productivity of agricultural lands, projects in construction, and electrical systems. High soil salinity in agriculture can cause a decrease in crop yields and hinder plant growth. Saline soils exert an osmotic pressure that is relatively higher, hence limiting the plants' water uptake. This thus results in poor agricultural yield with low-quality crops. Therefore, the management of soil salinity is of great importance for effective farming, particularly in arid and semi-arid zones, where the problems of salinity are more common. Soil salinity as a whole affects the general stability and durability of the structures. Such high levels of salinity could result in the corrosion of building materials and infrastructure, demanding that more resistant materials or protective measures be used. Soil salinity and resistivity should be highly determined when selecting the sites and designing the foundation for long-term stability purposes (Liangshan et al., 2010). In designing the earthing system in electrical engineering, soil salinity is a critical factor. The resistivity of soil, which is dependent on salinity, determines the effectiveness of grounding systems. Defective grounding may lead to electrical failures and endanger safety. In such a case, soil salinity and resistivity assessment become an essential exercise while designing electrical infrastructure and in this special case, substation and other electrical infrastructure (Hovhannissian et al., 2019).

Remote sensing is turning out to be very effective for soil salinity detection and mapping. It is also possible to assess soil salinity over extensive areas effectively and for a low cost, using satellite images and several indices, among them the Normalized Difference Salinity Index (NDSI). Such remote sensing technologies are able to find out the spatial differences of soil salinity, which is important for location-specific agricultural purposes and infrastructure planning. For example, from Landsat images, vital data can be collected for soil moisture, salinity, etc., which would help in making important decisions related to land and different construction projects for farmers and engineers, respectively (Shrestha, 2006, as cited in Sharma & Bhargawa, 1988; Dwivedi et al., 1999; Metternicht, 2003).

Soil resistivity is a very important parameter for designing earthing systems for electric installations. Depending on the salinity, moisture content, and composition, the resistivity of soil widely differs. High resistivity of soil may result in an improper grounding of the circuit that would increase the electrical faults or equipment failures. For example, a number of such studies in Sri Lanka have revealed that the earth resistance values in a number of substations exceed recommended levels, indicating poor grounding performance. Identification of various areas of soil resistivity will be resourceful to improve the reliability of electrical systems and safety in power distribution (Jayawardena & Mäkelä, 2021).

Based on the above-mentioned background, this study will develop a predictive model that uses machine learning to get further insights or predictions from the determination of soil salinity and resistivity. The model will thus bring about accurate prediction of soil salinity levels in the regions by integrating remote sensing information with measured soil properties. Therefore, predictions from this model can augment the current knowledge used to manage soil salinity and enhance agricultural practice, construction design, and electrical system reliability. Machine learning can be employed in the analysis of such complex data, very useful to achieve high levels of accuracy in the recommendations and interventions needed in the management of soil salinity (Liangshan et al., 2010). To sum up, soil salinity is a multifaceted problem that affects many other spheres. This will be one of the key areas where remote sensing and machine learning can be used in providing an innovative way that will detect, map, and manage soil salinity for more sustainable agricultural practice, safe construction methods, and reliable electrical systems.

Literature review

Importance of Earth Resistance and Resistivity for Electrical Distribution Networks

In the design and maintenance of electrical distribution networks, earth resistance and resistivity are critical parameters. The study of earth resistivity is essential for creating effective grounding systems that ensure safety and reliability in electrical and electronic systems. Research indicates that earth resistivity significantly influences the effectiveness of earthing systems and directly impacts the performance of electrical installations (Axis, 2021). Measuring earth resistivity allows engineers to optimize grounding systems, particularly in areas with high soil resistivity, where equipment failures and safety hazards are prominent. In the Sri Lankan context, the Ceylon Electricity Board reported that many substations had earth resistance values exceeding 20Ω, which is above the recommended levels, thereby affecting the performance and reliability of the local power grid.

Technological advancements such as Electrical Resistivity Tomography (ERT) and Electromagnetic Induction (EMI) have simplified the mapping of soil resistivity distribution. These methods facilitate the identification of areas with high resistivity, thereby improving equipment placement and grounding system designs (Hovhannissian et al., 2019).

Problems with Establishing Electrical Distribution Networks

One of the significant challenges in establishing electrical distribution networks globally and locally is ensuring the adequacy of earthing systems, particularly in areas with high soil resistivity. High soil resistivity can lead to inefficient grounding systems, which may result in electrical faults and equipment failures (IET, 2022). In regions with diverse soil properties, the cost of implementing a reliable network increases due to the necessity for more advanced earthing solutions (Axis Electricals, 2020). The difficulty in mapping soil resistivity across various terrains further complicates this issue. For instance, in Sri Lanka, elevated resistivity values have been linked to transformer failures in several regions (IET, 2022; LSP, 2022).

Environmental factors, including soil moisture and salinity, significantly impact the performance of earthing systems. In regions where soil properties fluctuate seasonally, such as the Niger Delta, the spatial variability in soil resistivity can adversely affect the longevity and reliability of electrical distribution infrastructure (Du et al., 2022). Furthermore, the cost and complexity associated with conducting ground-based resistivity measurements contribute to the high expenses and time requirements involved in deploying new infrastructure (IET, 2011; Axis Electricals, 2023). These challenges underscore the importance of understanding local soil conditions for effective earthing system design.

Significance of Geospatial Technologies with AI-Based Methods in Establishing Electrical Distribution Networks

The integration of geospatial technologies, such as Geographic Information Systems (GIS) and remote sensing, with artificial intelligence (AI)-based methods has significantly enhanced the efficiency and accuracy of establishing electrical distribution networks. GIS and remote sensing provides critical insights into spatial variations in soil properties, which are essential for designing cost-effective and efficient grounding systems (Bosisio et al., 2021). By leveraging satellite imagery and machine learning algorithms, such as random forest and Lasso regression, engineers can effectively predict optimal locations for earthing systems and optimize the placement of substations (Axis Electricals, 2023). This technological synergy allows for improved decision-making in the planning and implementation of electrical infrastructure.

AI-based models applied to geospatial data have demonstrated high accuracy in predicting soil resistivity, significantly reducing the need for extensive on-ground surveys. In Sri Lanka, a machine learning-based predictive model has been developed to map soil resistivity, enabling electrical engineers to optimize earthing system designs using real-time data (Aliakbar et al., 2023). These models enhance cost-efficiency by minimizing the necessity for physical resistivity measurements across large areas. Additionally, the application of deep learning techniques in spatial modeling of soil salinity has outperformed traditional methods, resulting in more accurate predictions and improved decision-making regarding the placement of electrical infrastructure (Bosisio et al., 2021).

GIS and RS have proven to be an affordable and effective tool for locating, mapping, and following salt-affected areas, their geographical and temporal changes; such areas demonstrate dynamic changes depending on the type of land use and weather conditions. Data collected through remote sensing, such as satellite imagery, can be combined with ground-based measurements to create electrical conductivity soil maps on a regional scale. Table 1 summarizes the 13 bands of the Sentinel-2 satellite that have been used in the present investigation and that are critical for creating indices such as NDVI and NDSI.

Table 1: Details of the 13 Bands of Sentinel-2

Band	**Central Wavelength (nm)**	**Resolution (m)**	**Description**
B01	443	60	Blue: Water quality assessment
B02	490	10	Green: Vegetation monitoring
B03	560	10	Red: Land cover classification
B04	665	10	Red edge: Vegetation stress detection

Band	Central Wavelength (nm)	Resolution (m)	Description
B05	705	20	Vegetation red edge: Vegetation monitoring
B06	740	20	Vegetation red edge: Vegetation monitoring
B07	783	20	Vegetation red edge: Vegetation monitoring
B08	842	10	Near-infrared: Vegetation, water assessment
B8A	865	20	Vegetation red edge: Vegetation monitoring
B09	945	60	Water vapor: Atmospheric correction
B10	1375	60	Cirrus: Cloud detection
B11	1610	20	SWIR: Moisture content, mineral identification
B12	2190	20	SWIR: Moisture content, mineral identification

Machine learning approaches have been better and higher in classification accuracy when looking through hyperspectral and multispectral remotely sensed data (Shafri, 2017). Machine learning may develop algorithms to learn from machine-readable data to model the ground-level abundance of airborne particulate matter while pinpointing global dust sources (David et al., 2017). Machine learning provides an excellent opportunity in the use of improving the analysis and interpretation of remote sensing data. For instance, with regard to predicting soil salinity, one can use a study that compares the performance of deep learning models with shallow machine learning models. The researchers applied the Random Forest and Support Vector Machine algorithms to multi-spectral Sentinel-2 data to obtain an overall accuracy for land-cover classification of the highest level (Singh et al., 2019). The results indicate better class accuracy of classification when the texture features were incorporated into the analysis, especially in the case of successional forests.

There have been various previous research on salinity associated with soil using remote sensing and GIS conducted by a number of researchers. Earth resistivity was predicted from Landsat 7 satellite imagery and field measurement-based data. The study yielded maps of earth resistivity wherein the confidence level was estimated at 85.71% (Norsangsri & Kulworawanichpong, 2009). Burnner et. al (2002) conducted a study on generating soil electrical conductivity maps at the regional level by integration of ground measurements and remote sensing data. This comprehensive study compared several methods of finding correlation with different outputs, keeping in consideration the effect of varied conditions such as variation in temperature and moisture content. The research concluded that the salinity maps obtained with the SCM algorithm un-calibrated showed a better correlation (Brunner et al., 2007).

A study has been carried out to analyze the soil resistivity and the effectiveness of designing grounding systems. This study achieved the measurement for soil resistance at two spots and hence analyzed the grounding system under both wet and dry soil conditions. This suggested that although looking at the data acquired from dry and wet conditions, it is not clear whether it was in fact dry or wet; hence, testing more locations would provide much more informative data (Malanda et al., 2018).

The spatial variability of soil resistivity for assessing corrosion risk and intensity was examined by Okiongbo et al., (2019). The corrosion risk and intensity in surface soils of four geomorphic zones in the Niger Delta were studied on the basis of measurements of spatial variation in soil resistivity. This property is related to the corrosion of ferrous metal. Sixty-five Schlumberger vertical electrical soundings, coupled with boring, up to a depth of about 5 meters, were carried out (Okiongbo et al., 2019).

Methodology

This research is illustrated with the use of different types of soil salinity analysis, like satellite imaging techniques, ground-based measurements, and machine-learning algorithms. The results obtained in these research works prove that the application of remote sensing and GIS for purposes of mapping and monitoring soil properties in the assessment of corrosion risk or the design of effective grounding systems can be very useful. In most studies, there were observations on the need for deep data collection and study toward improving the accuracy and reliability of the results.

In the current study, remote-sensed data from Sentinel-2 satellite was used to derive indices that will be used as input in machine learning models. It provided an effective approach for predicting soil resistivity over different locations within the area of Rathmalana. Figure 1 depicts the soil resistivity predictive model flow, indicating how remote sensing data and ML techniques were integrated.

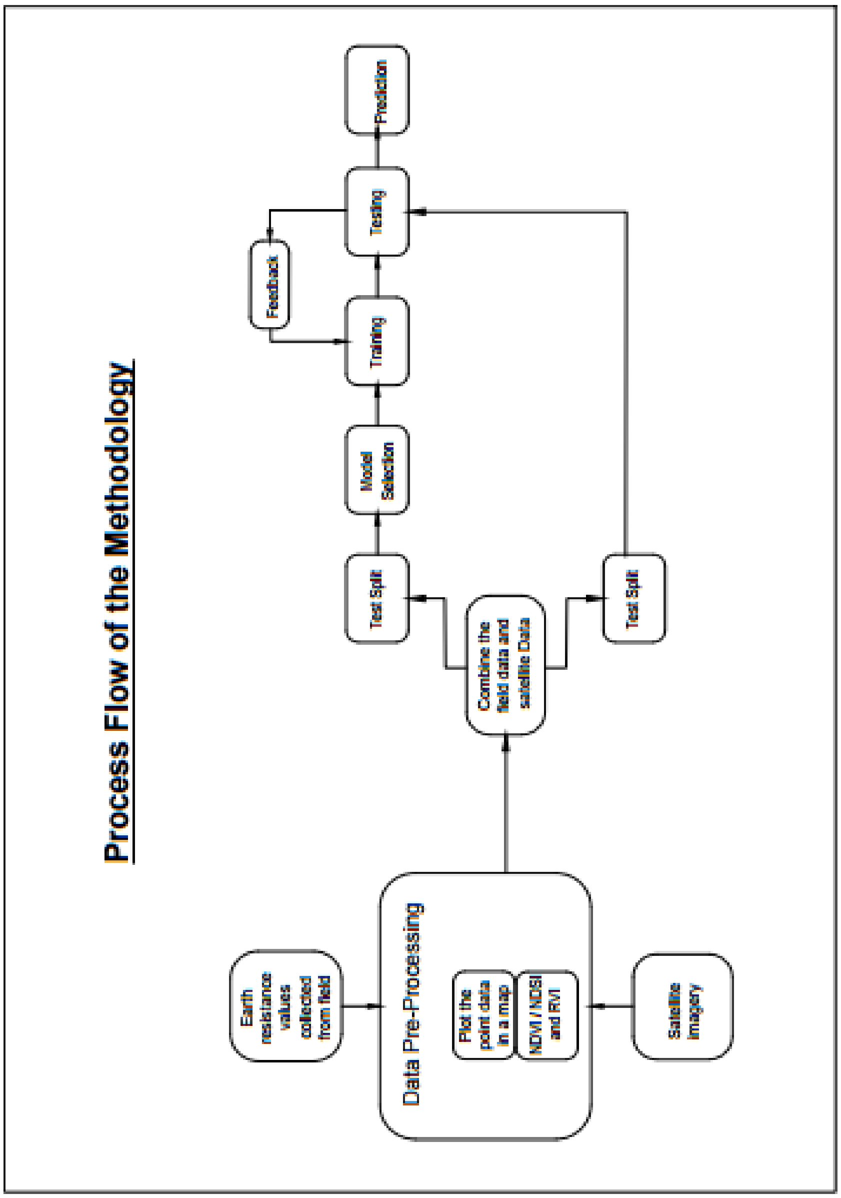
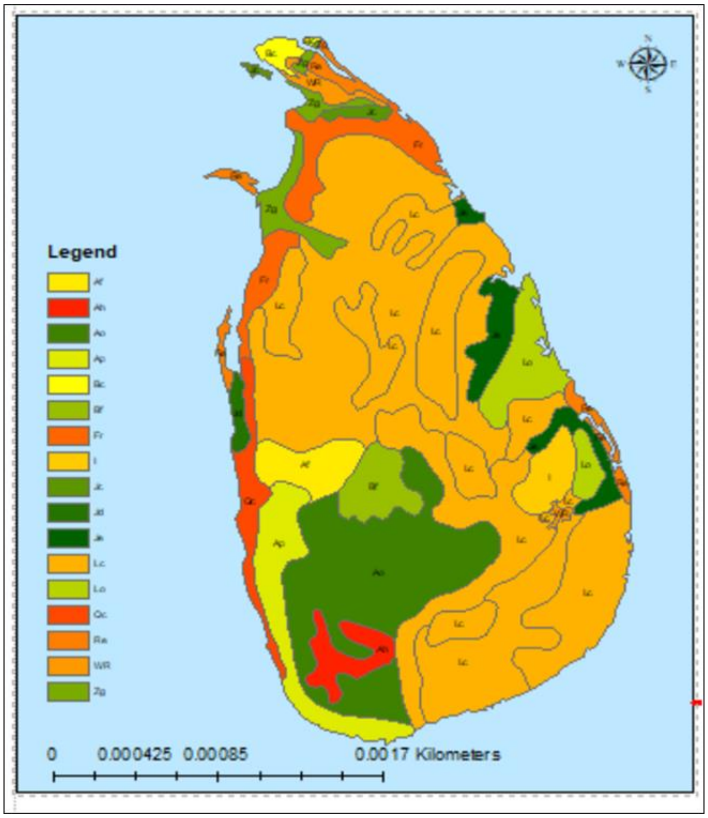


Figure 1: Process Flow of the Soil Resistivity Predictive Model



Source the digital soil map of the world (FAO, F. (1995). Digital soil map of the world-famous, Rome.)

Figure 2: Soil Map of Sri Lanka

The content, the agricultural practices that are conducted in the area, such as paddy, rubber, and coconut plantation, and even the existence of the marshland assist in influencing the soil properties and, therefore, cause variability in soil resistivity. Variation of soil types is shown in figure 2. The present study has also selected Rathmalana motivated by its high rate of transformer failure, which is an indicator of hidden contributions related to soil resistivity and salinity. Figure 3 gives the selected study area.

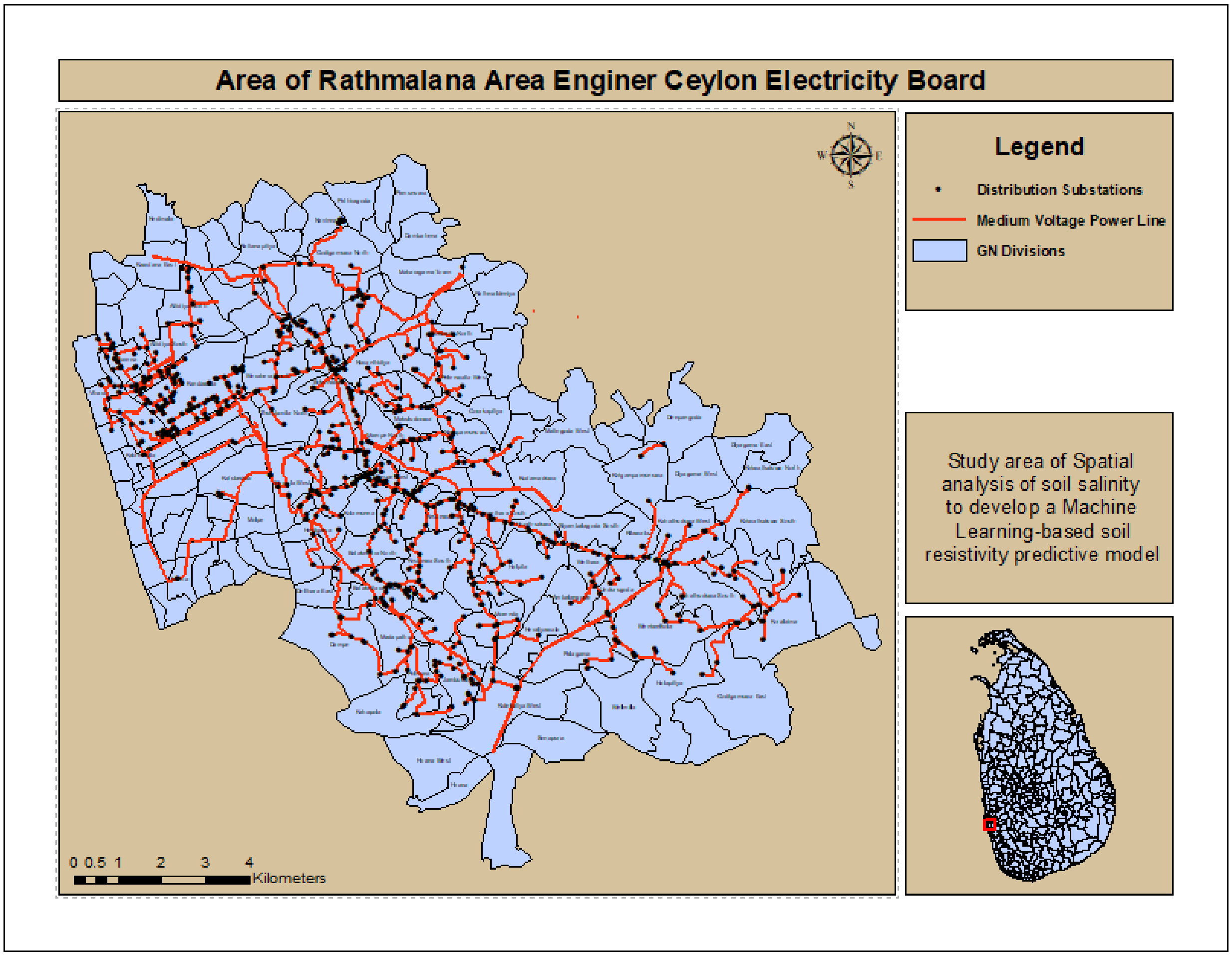


Figure 3: Study area

Both primary and secondary data were employed for the study. A brief description of data used for the study is described as follows.

Primary Data: Measurement of soil resistivity within the Rathmalana area was conducted through primary data collection at various locations. The required measurement was done using the Wenner four-pin method, a standard method to measure soil resistivity. From each field, 85 different locations were measured, ensuring that the distance among the points taken was more than 10 metres in order to avoid spatial correlation issues. This approach allowed for soil resistivity spatial variability to be easily appraised over different soil types.

Secondary Data: The secondary data was collected through remote sensing, especially Sentinel-2 satellite imageries. This data was very significant because it could be analyzed for the saltiness of the soil and related salinity parameters. The study generated different indices to check the level of salinity in soil, among them NDSI, NDVI, and RVI. The 13 spectral bands carried by the Sentinel-2 satellite are very useful for the sensing of different dimensions of land surfaces. These data were collected through the integration of primary and secondary data in a manner that the resultant dataset would be robust enough to carry out analysis. Priority was thus given to identifying soil salinity and resistivity spatial variability, relevant for electric installations grounding systems.

Table 2 provides an overview of the data types, sources, and methods of collection or processing used in this study.

Table 2: Data Description

Specific Objective	**Data Type**	**Data Source**	**Method of Collection/Processing**
Identify spatial variation of soil salinity	Secondary Data	Sentinel-2 Imagery	Calculation of NDVI, NDSI, RVI using satellite imagery
Identify spatial variation of soil resistivity	Primary Data	Ground Data	Field data collection using the Wenner four-pin method
Develop predictive model for earthing locations	Primary & Secondary	Ground & Satellite Data	Supervised classification and machine learning algorithms

The study was based on three machine learning models: Linear Regression, Lasso Regression, and Random Forest for the development of a predictive model of soil resistivity. The reason for the selection of these indexes is that they have been commonly used for the similar type of studies (Naimi et al., 2021)

Linear Regression:

It is a model that predicts a dependent variable, which is the resistivity of soil, against one or many independent variables like NDVI, NDSI, RVI to try to find an optimal linear relationship between the predictors and the target variable. The linear regression equation will be as: $y = b0 + b1x1 + b2x2 + \ldots + bnxn$

where y is the dependent variable and x1, x2, …, xn are independent variables, and b0, b1, …, bn are coefficients calculated by minimizing the sum of squared errors according to the least squares method on training data (James et al., 2013).

Lasso Regression:

L1 regularization is added into the cost function to prevent the overfitting behavior by including a penalty term. The mathematical form of the Lasso regression model is formulated in this equation:

$y = X\beta + e$

where y is the dependent variable, X is the independent variables, β is the vector of coefficients, and e is the error term. The cost function for the Lasso regression has the MSE within it, and it also has an L1 regularization term that helps to select features through reduction to zero in coefficients where features are not important (Tibshirani, 1996).

Random Forest:

It is also an ensemble of decision trees that can be used for classification or regression. It follows the process of training and building many decision trees. In a general aspect, it outputs the mean prediction of the individual trees. Each tree is built on a different bootstrap sample of observations, drawn from the data with random subsets that give uncorrelated trees. It is usually done at the level of splitting nodes, so it boosts classification.

Data Pre-processing

The collected data was subjected to a data preprocessing step prior to training the models, during which soil salinity and resistivity values were mapped and correlated with each other to form a complete dataset. The estimated NDVI, NDSI, and RVI are:

- NDVI = NIR-R/NIR+R = (Band 08 NIR – Band 04 Red)/ (Band 04 Red + Band 08 NIR)
- NDSI = SWIR-R/R+SWIR = Band 11 Red – Band 04 NIR / Band 04 Red + Band 11 NIR
- RVI = NIR/R = Band 08 NIR / Band 04 Red.

where NIR is the near-infrared band, R is the red band, and SWIR is the shortwave infrared band. Once these indices are calculated, the data was partitioned into subsets for training and testing. 70% of data was used in model building, and the remaining in validation. This split is important for the purpose of evaluation of the model performance and to maintain robustness in predicting soil resistivity based on salinity levels of the soil.

Figures 4, 5, and 6 show the spatial distribution maps for NDVI, NDSI, and RVI, respectively

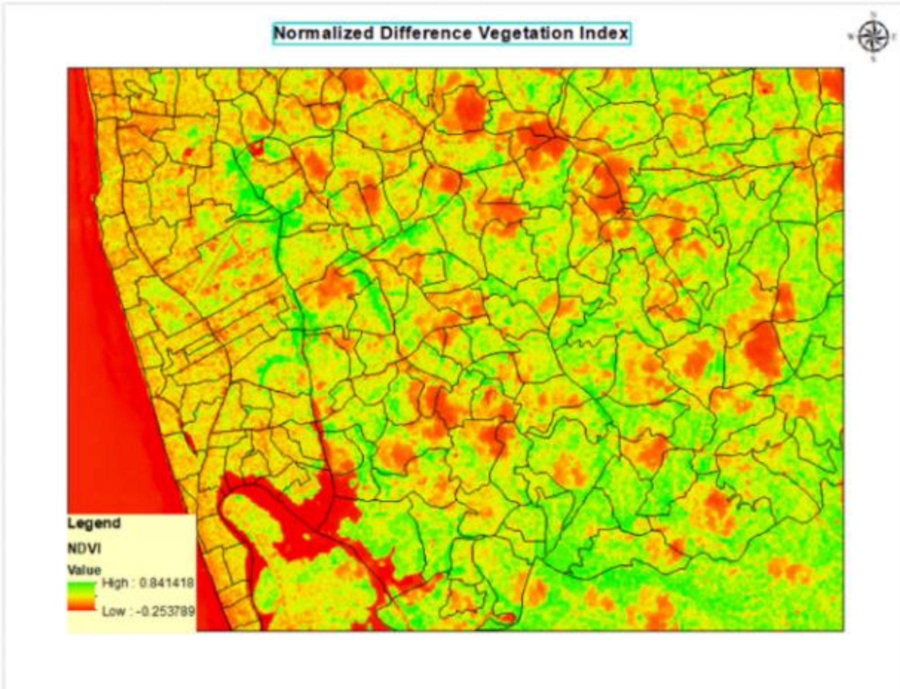


Figure 4: NDVI of the study area in 2022

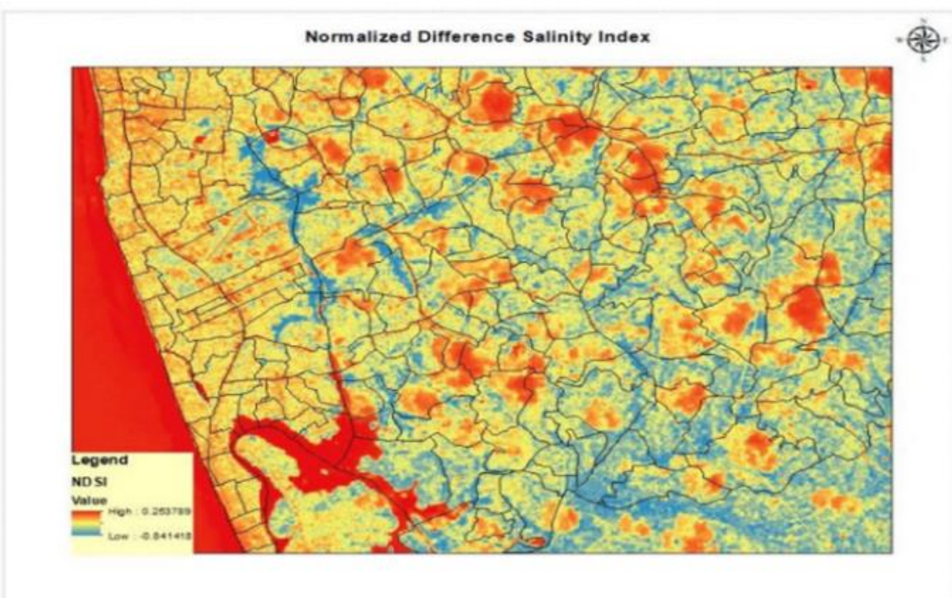


Figure 5: NDSI of the Study Area in 2022

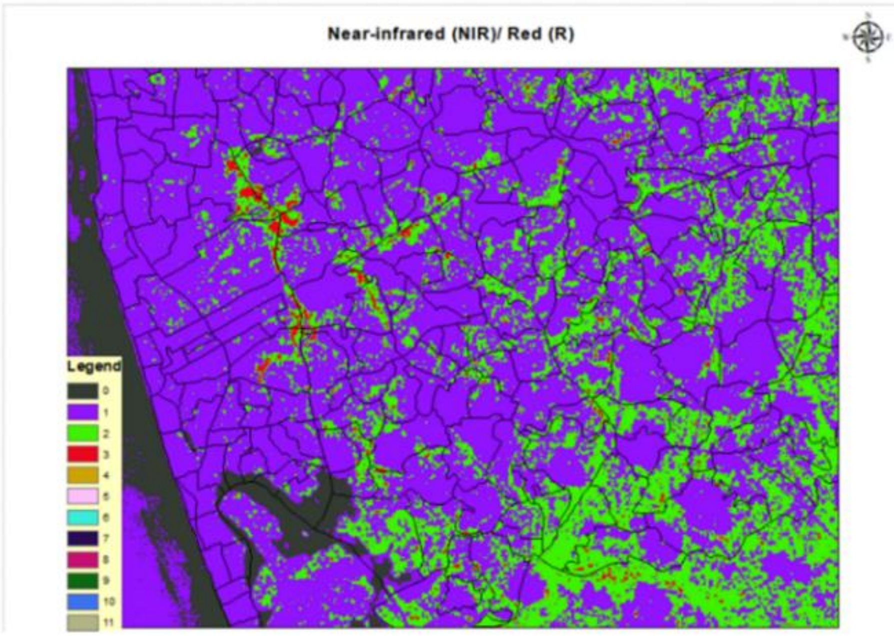


Figure 6: RVI of the Study Area in 2022

Model Training and Testing

During the training phase, the models were fit on the training data so that the algorithms could learn the relationships between the independent variables (NDVI, NDSI, RVI) and the dependent variable (soil resistivity). The models were then tested on reserved testing data to evaluate their predictive accuracy. Performance indicators like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were used to gauge the models' performance, providing insights into how well the models generalize to new data.

In essence, the research aimed to develop a predictive model using machine learning approaches for estimating soil resistivity based on primary field and secondary remote sensing data. These methodologies are expected to enhance the understanding of soil properties in the Rathmalana area, contributing to the design of safer and more reliable grounding systems for electrical installations.

Results and Discussion

The predictive models developed in this study—Linear Regression, Lasso Regression, and Random Forest—were evaluated using various metrics to assess their performance in predicting soil resistivity based on soil salinity levels. The results of each model are detailed in the following sections.

Linear Regression

The model was trained using the collected data, with soil resistivity as the dependent variable and NDVI, NDSI, and RVI as independent variables. The model's performance was evaluated using RMSE and R-squared values. RMSE measures the average magnitude of the errors in the same units as the original dependent variable, while R-squared indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. The Linear Regression model achieved an RMSE of 0.78 and an R-squared value of 0.65, suggesting a moderate fit between the predicted and actual soil resistivity values.

Lasso Regression

Lasso Regression, the second predictive model, incorporated L1 regularization to prevent overfitting and perform feature selection. The model's performance was evaluated using the same metrics as the Linear Regression model. The Lasso Regression model achieved an RMSE of 0.72 and an R-squared value of 0.71, indicating a better fit compared to the Linear Regression model. The L1 regularization in Lasso Regression helped reduce the impact of less significant variables, leading to improved model performance.

Random Forest

Random Forest, an ensemble of decision trees, was the third predictive model. This model is known for handling non-linear relationships and providing robust predictions. The Random Forest model achieved an RMSE of 0.65 and an R-squared value of 0.77, outperforming both the Linear Regression and Lasso Regression models. The ensemble nature of Random Forest, along with its ability to capture complex patterns in the data, contributed to its superior performance in predicting soil resistivity. Table 3 summarizes the performance metrics for each predictive model:

Table 3: Model Performance Metrics

Model	RMSE	R-squared
Linear Regression	0.78	0.65
Lasso Regression	0.72	0.71
Random Forest	0.65	0.77

Figure 7. Comparison of predicted resistivity soil values of the best model, Random Forest model with actual values. A scatter plot is showing predictions are closer by the Random Forest model, but not exact; it has discrepancies mostly in areas with extreme resistivity values.

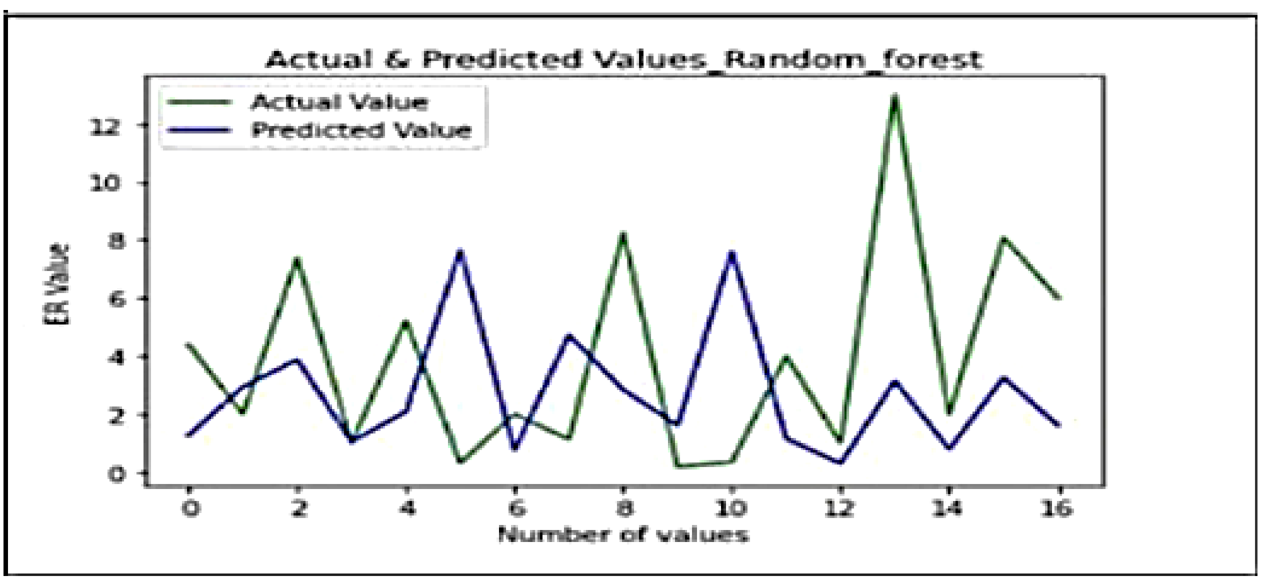


Figure 7: Actual vs. Predicted Soil Resistivity Using Random Forest Model

Discussion

The findings of this study highlight the significant relationship between soil salinity and resistivity in the Rathmalana area. The predictive models developed using machine learning techniques, particularly Random Forest, have shown promising results in estimating soil resistivity based on soil salinity levels. The superior performance of the Random Forest model can be attributed to its ability to handle non-linear relationships and its ensemble nature, which reduces the impact of individual model weaknesses. The model's ability to capture complex patterns in the data and provide robust predictions makes it a valuable tool for estimating soil resistivity in areas with varying soil salinity levels.

The implications of this study are significant for the electrical engineering sector. Accurate prediction of soil resistivity is crucial for designing effective grounding systems that ensure the safety and reliability of electrical installations. By incorporating soil salinity data obtained from remote sensing techniques, such as satellite imagery, and using machine learning models like Random Forest, electrical engineers can make informed decisions about the placement and design of grounding systems. However, it is important to note that the accuracy of the predictive models is dependent on the quality and quantity of the input data. While this study has demonstrated the potential of machine learning in predicting soil resistivity, further research is needed to validate the models in different geographical regions and under varying soil conditions.

In conclusion, this study has successfully developed a machine learning-based predictive model for soil resistivity, with the Random Forest model showing the best performance. The integration of soil salinity data obtained from remote sensing and ground-based measurements has proven to be a valuable approach in estimating soil resistivity. The findings of this study can contribute to the design of safer and more reliable grounding systems in electrical installations, ultimately enhancing the overall safety and reliability of the electrical grid.

Interpretation of Model Results

The results of the machine learning models indicate that while there is a discernible relationship between soil salinity indices and soil resistivity, the models were not able to fully capture this relationship. The Random Forest model, despite being the best performer, still showed significant errors, particularly in predicting extreme values. This suggests that the relationship between soil salinity and resistivity is complex and likely influenced by additional factors not captured in this study. The Linear Regression and Lasso Regression models, both of which assume a more straightforward relationship between variables, were less effective in predicting soil resistivity. This highlights the importance of considering non-linear relationships in environmental modeling, particularly when dealing with heterogeneous data such as soil properties.

Implications for Practical Applications

The findings have important implications for the design and placement of earthing systems in electrical and telecommunication infrastructure. Accurate prediction of soil resistivity is crucial for selecting optimal locations for these systems, as high resistivity can lead to poor grounding and increased risk of equipment failure. The relatively high error rates in the models suggest that while machine learning provides a valuable tool for this purpose, further refinement is needed. The study also underscores the potential of integrating remote sensing with machine learning to create predictive models for environmental parameters. While the results were promising, they also point to the need for incorporating additional variables—such as soil moisture content, organic matter, and temperature—to improve model accuracy.

Limitations

One of the limitations of this study was the reliance on a limited set of remote sensing indices (NDVI, NDSI, RVI) as predictors. While these indices are useful for assessing vegetation and salinity, they may not fully capture all factors influencing soil resistivity. Future studies should explore the inclusion of other variables, such as soil temperature, moisture content, and soil texture, to enhance predictive accuracy. Additionally, the models were trained and tested on data from a specific geographical area (Rathmalana). The generalizability of these models to other regions with different soil types and climatic conditions remains to be tested. It is recommended that similar studies be conducted in other regions to validate the findings and refine the models. Finally, while the Random Forest model showed the most promise, further experimentation with hyperparameter tuning and the incorporation of ensemble methods could potentially improve model performance.

Conclusion and Recommendations

Conclusion

This study set out to develop a machine learning-based predictive model for soil resistivity using remote sensing data, specifically focusing on soil salinity indices in the Rathmalana area of Sri Lanka. The research demonstrated the potential of integrating remote sensing and machine learning techniques to predict soil resistivity, a critical factor in the design and placement of earthing systems for electrical and telecommunication infrastructure. The three machine learning models—Linear Regression, Lasso Regression, and Random Forest—were tested, with the Random Forest model outperforming the others in terms of accuracy. However, despite its superior performance, the Random Forest model still exhibited significant error margins, particularly in predicting extreme resistivity values. This suggests that while the relationship between soil salinity and resistivity is evident, it is more complex than initially assumed, likely influenced by additional environmental and soil variables not included in this study. The study contributes valuable insights into the application of machine learning for environmental modeling, highlighting both the strengths and limitations of current approaches. It also underscores the importance of selecting appropriate models and variables when dealing with complex, heterogeneous data like soil properties.

Recommendations

Based on the findings, several recommendations can be made for future research and practical applications:

- Incorporate Additional Variables**:** Future studies should consider incorporating other environmental variables such as soil moisture content, organic matter, soil temperature, and texture to enhance the predictive accuracy of soil resistivity models. These factors are known to influence soil resistivity and could help in developing more robust models.
- Expand the Study Area**:** To improve the generalizability of the findings, similar studies should be conducted in different geographical regions with varying soil types and climatic conditions. This will help in validating the models and ensuring they are applicable in a wider range of environments.

- Explore Advanced Machine Learning Techniques**:** While the Random Forest model showed promise, further exploration of advanced machine learning techniques such as boosting algorithms (e.g., XGBoost) or deep learning models could lead to improved performance. Additionally, hyperparameter tuning and cross-validation should be employed to optimize model settings.
- Practical Applications for Infrastructure Design**:** The insights gained from this study should be applied to the design and placement of earthing systems in electrical and telecommunication infrastructure. By using predictive models to identify optimal locations with lower soil resistivity, the reliability and safety of these systems can be significantly enhanced.
- Continuous Monitoring and Model Updating**:** Given that soil properties can change over time due to environmental factors, it is recommended that continuous monitoring be implemented. The predictive models should be regularly updated with new data to maintain their accuracy and relevance.
- Collaboration Between Disciplines**:** The study demonstrates the importance of interdisciplinary collaboration between geographers, data scientists, and engineers. Such collaborations can lead to more comprehensive models and solutions that address complex environmental challenges effectively.

This research provides a foundational step towards leveraging machine learning and remote sensing in environmental modeling, particularly for applications in infrastructure planning and design. While there is still much to be explored and refined, the integration of these technologies offers significant potential for improving the way we understand and manage the natural environment.

Reference

Aliakbar, A., Jafari, M., & Moghaddam, M. (2023). AI and machine learning for soil analysis: An assessment of methods and applications. Bioresources and Bioprocessing, 10(1), 1-15. https://doi.org/10.1186/s40643-023-00710-y

Axis Electricals. (2020). Understanding the significance of soil resistivity in earthing system design. Retrieved from https://axis-india.com/importance-soil-resistivity-earthing-systems

Axis. (2021). What is earth resistance? How to measure it? [Blog post]. Axis-India. https://axis-india.com/what-is-earth-resistance-how-to-measure-it/

Axis Electricals. (2023). Understanding the significance of soil resistivity in earthing system design. Retrieved from https://axis-india.com/importance-soil-resistivity-earthing-systems

Bosisio, A., Berizzi, A., Morotti, A., Pegoiani, A., Greco, B., & Iannarelli, G. (2021). Machine learning and GIS approach for electrical load assessment to increase distribution networks resilience. Energies, 14(14), 4133. https://doi.org/10.3390/en14144133

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brunner, P. H. T. L., Li, H. T., Kinzelbach, W., & Li, W. P. (2007). Generating soil electrical conductivity maps at the regional level by integrating measurements on the ground and remote sensing data. International Journal of Remote Sensing, 28(15), 3341-3361.

David, C., Ghosh, S., & Gupta, A. (2017). Machine learning algorithms for modeling airborne particulate matter and identifying global dust sources. Environmental Science and Pollution Research, 24(14), 12345-12356. https://doi.org/10.1007/s11356-017-0168-2

Du, R., Chen, J., Zhang, Z., Chen, Y., He, Y., & Yin, H. (2022). Simultaneous estimation of surface soil moisture and salinity during irrigation with the moisture-salinity-dependent spectral response model. Water Science and Technology, 85(5), 1234-1246. https://doi.org/10.1016/j.pce.2022.05.001

FAO. (1995). Digital soil map of the world. FAO.

Hovhannissian, G., Podwojewski, P., Le Troquer, Y., Mthimkhulu, S., & Van Antwerpen, R. (2019). Mapping spatial distribution of soil properties using electrical resistivity on a long-term sugarcane trial in South Africa. Geoderma, 349, 56-67.

IET. (2011). Earthing systems: Principles and practices. Retrieved from https://electrical.theiet.org/media/1549/earthing.pdf

IET. (2022). Earthing - IET Electrical. Retrieved from https://electrical.theiet.org/media/1549/earthing.pdf

Indraja, P., Kumar, H. H., Mani, A., & Edukondalu, L. (2018). Mapping of soil salinity using remote sensing and GIS. The Andhra Agricultural Journal, 54, 306-312.

LSP. (2022). Earthing systems TNC TNS TNCS TT IT - Surge protective device - LSP. Retrieved from https://lsp.global/earthing-systems

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: With applications in R. Springer.

Jayawardena, I. M., & Mäkelä, A. (2021). Spatial and temporal variability of lightning activity in Sri Lanka. In Multi-Hazard Early Warning and Disaster Risks (pp. 573–586). Springer.

Liangshan, L., Zhang, J., & Wang, H. (2010). The influence of soil salinity on the corrosion of building materials and infrastructure. Journal of Environmental Management, 91(3), 589–594. https://doi.org/10.1016/j.jenvman.2009.10.017

Malanda, S.C., Davidson, I.E., Singh, E., & Buraimoh, E. (2018). Analysis of soil resistivity and its impact on grounding systems design. In 2018 IEEE PES/IAS PowerAfrica (pp. 324–329). IEEE.

Naimi, S., Ayoubi S., Zeraatpisheh M., & Dematte J.A.M.(2021). Ground observations and environmental covariates integration for mapping of soil salinity: a machine learning-based approach.Remote Sensing, 13(23), 4825.

Norsangsri K., & Kulworawanichpong T.(2009). Earth resistivity estimation based on satellite imaging techniques.WSEAS Transactions on Systems, 8(9),1061–1071.

Okiongbo K.S., Akpofure E.A.R.C.(2019). Spatial variability of soil resistivity for assessing corrosion risk and intensity in the Niger Delta.International Journal of Applied Environmental Sciences,14(1),37–51.

Shafri H.Z.(2017). Machine learning in hyperspectral and multispectral remote sensing data analysis.InArtificial Intelligence Science and Technology: Proceedings of the 2016 International Conference (AIST2016) (pp.3–9).

Shrestha R.P.(2006). Relating soil electrical conductivity to remote sensing and other soil properties for assessing soil salinity in northeast Thailand.Land Degradation & Development,17(6),677–689.

Singh A., Kumar S.(2019). Performance comparison of deep learning and shallow machine learning models for soil salinity prediction using Sentinel-2 data.Remote Sensing,11(12),1423.

https://doi.org/10.3390/rs11121423

Tibshirani R.(1996). Regression shrinkage and selection via the lasso.*Journal of the Royal Statistical Society: Series B(Methodological)*58(1),267–288.

https://doi.org/10.1111/j.2517–6161 .1996.tb02080.x