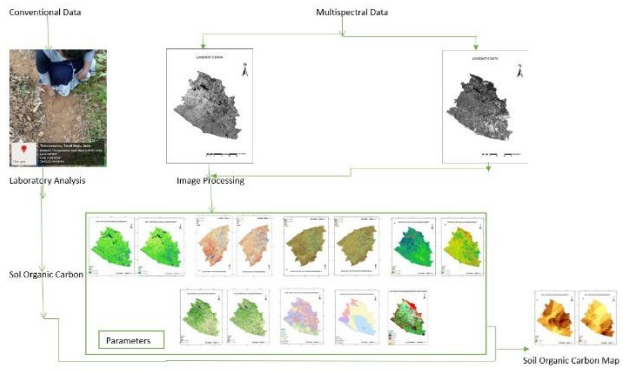
Global NEST Journal, Vol 27, No X, pp 1-11

Machine learning based random forest algorithm for prediction of soil organic carbon decompositions

Arthi Arul Pushpa[1*] and Satish Kumar Jeyapaul[2]
[1]Research Scholar, Department of Civil Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India,603203
[2]Associate professor, Department of Civil Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India, 603203
Received: 16/10/2024, Accepted: 01/05/2025, Available online: 20/05/2025
*to whom all correspondence should be addressed: e-mail: arthiap36@gmail.com
https://doi.org/10.30955/gnj.06890

Graphical abstract



Abstract

The Kanyakumari District was selected as a study area due to its diverse flora and favorable climate and soil conditions. Various remote sensing parameters were derived from Landsat 8 and 9 satellite data, including NDVI, BSI, NDMI, NDWI, and SAVI. Moreover, data on land use and land cover (LULC), geology, and soil type were considered. These parameters were utilized as inputs for Random Forest Regression Analysis to establish the relationship between Soil Organic Carbon (SOC) and each parameter, enabling predicting of future SOC levels. Comparing Landsat 8 and 9 data, BSI and NDWI displayed high correlations, while NDVI, NDMI, and SAVI exhibited medium correlations. To validate SOC predictions, 115 soil samples were collected from the field, and laboratory SOC content analysis was performed. Machine learning algorithms, specifically Random Forest Regression, were employed to predict SOC values. The predicted SOC values indicated spatial variations, with residential areas exhibiting low SOC and forested areas showing higher SOC due to minimal human disturbances. Creating a SOC map is instrumental in identifying regions requiring soil restoration and yield enhancement. This study underscores the utility of SOC mapping in guiding soil restoration efforts and enhancing agricultural productivity, with implications for precision farming and sustainable land management.

Keywords: SOC, BSI, NDVI, NDWI, NDMI, SAVI

1. Introduction

The soil exhibits variability across fields due to its diverse chemical and physical properties and its mixture of organic and inorganic constituents. SOC is a key component of the carbon cycle, significantly influences the soil ecosystem through its biological, physical, and chemical interactions. For soil quality to be maintained, soil organic matter, or SOM, is essential. It serves as the primary energy source for microorganisms and a vital regulator of soil structure and ecosystem productivity, contributing to the health of terrestrial ecosystems (Abdel-Kader, 2011).

SOC is beneficial for the health and productivity of Soil and provides a significant source of plant nutrients, develops the structure of Soil, enhances the capacity to store water, and provides a soil habitat organism (Alhameid, *et al*. 2017; Coming, 2014). By capturing and preserving SOCs in the Soil, we can also improve soil quality and productivity for food production and reduce the release of greenhouse gases. Carbon is the main component of SOC and helps give the Soil its water-holding capacity, soil structure, and determining the SOC is helpful in precision farming and essential for site-specific crop management (Franke and Menz, 2007).

SOC serves as a reservoir for carbon storage in terrestrial ecosystems. By reducing atmospheric carbon dioxide (CO_2) levels, carbon sequestration facilitated by higher SOC helps mitigate climate change. It acts as a long-term carbon sink, slowing down the rate of global warming (Qiu *et al*. 2021).

SOC is a vital component of the nutrient cycle within ecosystems. It acts as a source of nutrients such as nitrogen, phosphorus, and sulfur. It offers a microbial activity substrate, which breaks down organic matter and releases essential nutrients for plant growth. Adequate

SOC levels promote soil fertility, supporting healthy plant growth and ecosystem productivity (Basset *et al*. 2023; Villarino, 2021). SOC provides habitats and energy sources for diverse soil organisms, including microorganisms, fungi, insects, and earthworms. These organisms contribute to soil biodiversity, nutrient cycling, and decomposition processes. A healthy soil ecosystem supported by sufficient SOC levels fosters biodiversity, sustains beneficial organisms, and promotes various ecosystem services. SOC affects the water-holding capacity of soils. Organic matter can absorb and retain water, enhancing soil moisture content and reducing water runoff, improving water availability for plants, helping regulate hydrological cycles, and reducing the risk of soil erosion and sedimentation in water bodies (Bhunia *et al*. 2019).

SOC plays a crucial part in creating and preserving the structure of soil. Soil particles are bound together by organic substances encouraging aggregation and improving soil stability. Well-structured soils with sufficient SOC display enhanced porosity, permeability, and water infiltration capacity (Wright and Upadhyaya, 1998). These improvements support better root penetration, reduce soil compaction, and enhance overall soil health.

SOC is critical in soil resilience against environmental disturbances like droughts, floods, or land degradation. Soils with higher SOC levels exhibit better resistance and recovery mechanisms, retaining moisture during dry periods and maintaining nutrient availability. Sustainable land management practices that enhance SOC levels can improve soil resilience and contribute to ecosystem sustainability (Burrough *et al*. 1997). Topography and services are factors that affect the spatial distribution of soil properties in sustainable land development. SOM management helps maintain soil fertility, which is helpful in sustainable agriculture (Tisdall and OADES, 1982). SOC is generally between 0.5 – 4.0 % in dry land agricultural soils globally. SOC is an essential characteristic for assessing the quality of Soil and also a predominant parameter of agricultural soils and should be systematically observed.

Conventional soil surveys record soil properties at representative locations. In soil science, vegetation type, and condition are the most important as they reflect and modify land surface processes such as energy and mass transfer modelling (Clerici *et al*. 2016; Cheng *et al*. 2012). Of course, soil properties are highly adjustable, and accurate evaluation of the properties of Soil must consider the changeability. A new kriging technique was widely used in land resource inventory as a required spatial interpolation method in 1970 (Gilabert *et al*. 2002). Conventional methods of soil interpretation and analysis are cumbersome and take more time. The Walkley-Black (WB) approach is widely used to determine SOC, but little attention has been paid to its reliability and usefulness compared to the dry combustion method. Numerous studies have been conducted to find SOC in Soil. Remote Sensing can analyze and accurately determine the Soil's Organic Carbon content (Patel, 2018).

Remote sensing (RS) is finding and keeping track of Earth's physical properties by identifying radiation that has been backscattered and reflected in a distant area, typically on a satellite or airplane (Vijayakumar, M. and Ahilan, A., 2024). Unique cameras and sensors collect remote sensing or satellite images, allowing users to "feel" the Earth. The powerful Method for mapping soil properties, such as SOC in Remote sensing, is a vital attribute of the quality of Soil. A remote sensing spectral index based on short-wave infrared (SWIR) or near-infrared (NIR) wavelengths is used to quantify the spatial pattern of SOC. It is also used in SOC (Huang *et al*. 2021; Ismail and Yacoub, 2012) to predict nutrient availability in plants and the optimal condition of Soil using different spectral indices (Jaber *et al*. 2011). These indices involve reflectance at various spectral wavelengths as mathematical transformations.

Land use type, agricultural intensity, and fertilizer source are important factors that change soil properties (Kumar *et al*. 2020; Lamichhane *et al*. 2019). The permanent soil properties, such as soil pH and texture, and climatic environment can also affect the changes in the properties of Soil. SOC controls soil properties like color and nutrient retention capacity and helps soil structure improvement (Liu *et al*. 2011). In India, most soil maps are produced using traditional methods, and work was done by spatial techniques for prediction (Marchetti *et al*. 2012).

Landsats 9 and 8 are two essential satellite missions that have contributed significantly to our understanding of Earth's surface and dynamics. Landsat 9 and 8 offer valuable Remote Sensing data for studying SOC levels, allowing scientists and researchers to monitor and evaluate soil health and related environmental processes (Kavitha, P., *et al*., 2025). These satellite missions provide multispectral and thermal data, enabling the assessment of land surface characteristics, vegetation dynamics, and land cover changes—parameters closely linked to SOC (Masek, 2020; Mondal and Chakraborty, 2022). The multispectral capabilities of Landsat 9 and 8 allow for the identification and classification comprises many kinds of land cover, such as built-up areas, aquatic bodies, and vegetation. By analyzing the spectral responses in different bands, researchers can extract valuable information about the presence and distribution of SOC across landscapes (Oldfield *et al*., 2019; Ruirui and Xiaoting, 2022). By providing high-quality multispectral and thermal data, these satellite missions have enabled researchers to study soil health, carbon sequestration, climate change mitigation, water management, and ecosystem sustainability.

The machine learning (Ahilan, A., *et al*., 2023) algorithm creates statistical models and algorithms that enable computers to learn, predict the future, or make decisions without explicit programming. It is a subfield of artificial intelligence (AI) (Sundarasekar, R. and Appathurai, A., 2022). It entails the research of mathematical formulas and statistical models that allow computers to carry out operations or anticipate future outcomes using patterns and inferences drawn from data (Jamalabad *et al*. 2004;

Somasundaram *et al*. 2018). A popular machine learning method, Random Forest, can forecast SOC levels based on various input data (A. Ashvanth Louison and Ben Sujitha, 2024; Zeng, *et al*. 2010).

The methodology in this study is to find parameters such as NDVI (Normalized Vegetation Index), BSI (Bare Soil Index), NDMI (Normalized Moisture Index), NDWI (Normalized Wetness Index), and SAVI (Soil Adjusted Vegetation Index) using Landsat 8 and 9 Satellite data. The soil type, geology, and LULC are also considered in this study. The Random Forest Regression analysis is then used to find SOC and identify relationships between each parameter. Then, the SOC identified using Landsat 8 and 9 multispectral data is compared.

1.1. Hypothesis

- How does the integration of satellite-derived spectral indices (NDVI, BSI, NDMI, NDWI, SAVI) enhance the accuracy of SOC prediction compared to traditional soil sampling methods?
- How does Landsat 9 data, with its improved spectral resolution, contribute to better SOC prediction compared to Landsat 8?
- How does the application of the Random Forest Regression model improve the robustness and reliability of SOC mapping?
- How do variations in land use, land cover, and soil type affect SOC prediction in different regions of the Kanyakumari district?

2. Materials and methods

2.1. Study area

Top of Form

Bottom of Form

The Kanyakumari district is located between 8°03' and 8°35' north latitude and 77°15' and 77°36' east longitude. Kanyakumari district is bounded by Tirunelveli district to the North and northeast, Mannar Bay to the east, the Indian Ocean to the south, the Arabian Sea to the west, and Thiruvananthapuram district (Kerala) to the west. The Kanyakumari has an area of 1,672 km2.

Kanyakumari District has two regions, namely Edai Nadu and Nanjil Nadu. The Edai Nadu region includes Thovalai and Agastheeswaram taluks are part of the Nanjil Nadu region, Vilavancode and Kalkulam taluks are part of the Western Ghats. The Aralvaimozhy Pass separates these two regions, and these areas are separated by the Vaezhimalai (Vaezhi Hills). The topography of the Kanyakumari district includes the Western Ghats on the west coast and the sea on three sides northern border or the other side surrounded by land. The year-round warm weather ranges from 30°C (86°F) to 33°C (92°F). The Study Area Map of Kanyakumari District shown in **Figure 1**.

2.2. Soil sample collection

Soil samples were collected at random depths of 0 cm and 20 cm from various LULC categories for validation. The latitude and longitude of 115 soil samples were recorded using a handheld GPS, as shown in **Figure 2**. SOC analysis of these samples was conducted in the laboratory using the Walkley-Black (WB) wet oxidation method (Liu *et al*. 2011).

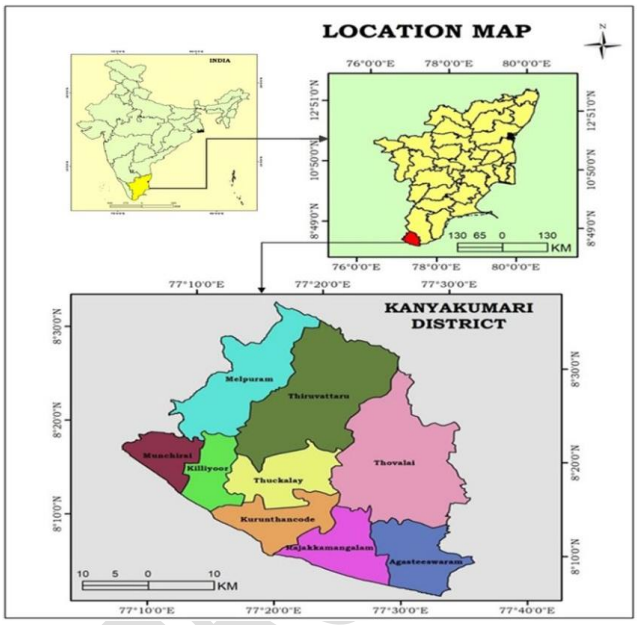


Figure 1. Study Area Map of Kanyakumari District

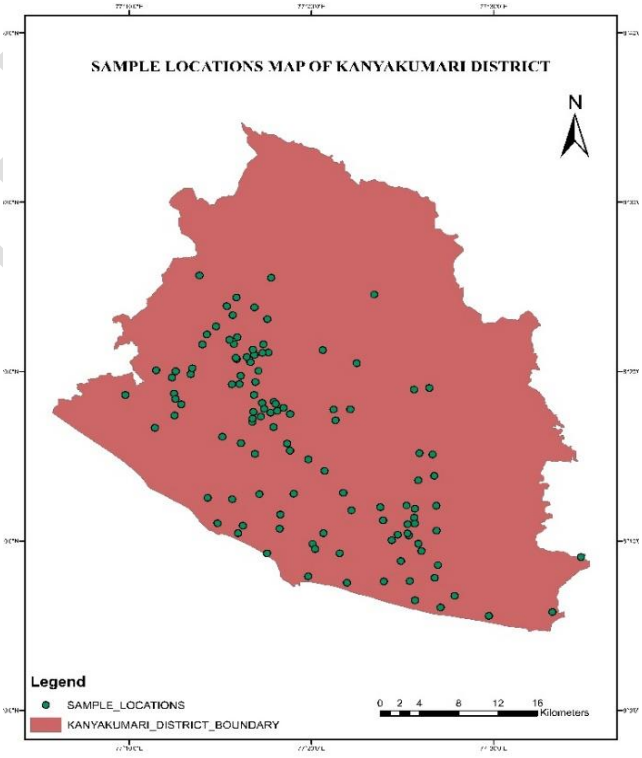


Figure 2. Sample Locations of the Soil Samples

2.3. Data collection

The SOC training/validation dataset was collected from the field in March 2023, and organic carbon measurements were collected through field studies using the Random Sampling Method. Landsat 8 and 9 data were used to predict SOC retrieved from USGS Earth Explorer. The Landsat 8 and 9 are Earth-observing satellites launched by The USGS and NASA. In 2013, Landsat 8 was launched. while Landsat 9 was scheduled for a 2021 launch (based on information available as of September 2021). Both missions are dedicated to Earth observation

and land imaging, designed with devices such as the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS), which enable the acquisition of multispectral pictures for a variety of applications. Landsat 9 aimed to provide enhanced data quality and continuity with previous Landsat missions, contributing to long-term, consistent Earth observation data that is freely accessible to the public and beneficial for uses like disaster management, forestry, land use planning, agriculture, and environmental monitoring.

3. Methodology

As mentioned in **Figure 3**, the data is collected using Multispectral (Landsat 8 and 9) and conventional Methods. The conventional method collects the 0 to 20 cm topsoil using random sampling. Then, the soil is sieved, taken to the Laboratory, and oven-dried for 24 hours. Then, by using the Walkley black method, the SOC is identified. Then, using the Multispectral Method, the Landsat 8 and 9 satellite images were preprocessed, and parameters like NDVI, BSI, NDWI, NDMI, and SAVI were identified, and the relationships between them were analyzed. Then, the SOC is predicted using Random Forest regression analysis, and a digital SOC Map is prepared.

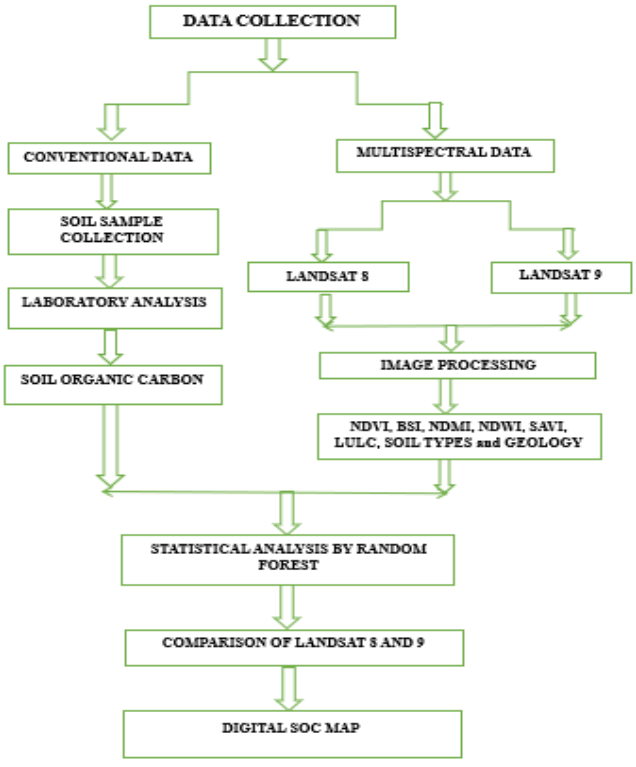


Figure 3. Methodology

3.1. Parameters

3.1.1. Bare soil index (Bsi)

The ground not covered with grass, other habitat cover, wood chips, gravel or rock surfaces, or artificial coverings is called Bare Soil. The BSI index has a numerical value associating red, near-infrared, green, and blue bands to identify soil changes. BSI's value depends on the soil background's evolution and sun angle. This is precise about Soil darkening as it is about vegetation growth. The Soil values of BSI are used to lower the difference between sun and shade soils and minimize dry and wet soil conditions. BSI is calculated according to existing using the following formula:

$$BSI = ((\mathbf{Red} + \mathbf{SWIR}) - (\mathbf{NIR} + \mathbf{Blue})) / ((\mathbf{Red} + \mathbf{SWIR}) + (\mathbf{NIR} + \mathbf{Blue}))$$

3.1.2. Normalized Differential Vegetation Index (Ndvi)

Normalized Difference Vegetation Index (NDVI) values, a metric for identifying vegetation, are estimated using the TM data's red channel (band 3) and NIR (band 4) status and monitoring vegetation change. NDVI is calculated by the formula.

$$\mathbf{NDVI} = (\mathbf{NIR} - \mathbf{Red}) \ / \ (\mathbf{NIR} + \mathbf{Red})$$

NDVI values range from -1 to +1, with the (+1) value (high) denoting vegetation or high backscatter surfaces and the (-1) value denoting areas devoid of vegetation or low backscatter. Negative (low) values indicate the presence of water, clouds, Bare soils, and non-healthy vegetation, which have equal or nearly equal backscatter for NIR and Red region, resulting in lower values. Healthy green foliage has a high reflectance of NIR radiation and, therefore, a high value. NDVI obtained from Landsat 8, 9 TM data was used to predict SOC. NDVI is generally sensitive to plant biomass and nitrogen status.

3.1.3. Normalized differential moisture index (Ndmi)

The NDMI is a number between -1 and +1, where the lowest values indicate low vegetation moisture content and the highest numbers signify a significant level of moisture. A decrease in NDMI shows water stress, and an abnormally high NDMI value indicates flooding. NDMI is sensitive to Moisture level of vegetation. The positive values indicate a high level of vegetation moisture content and vice versa.

$$\mathbf{NDMI} = (\mathbf{NIR} - \mathbf{SWIR}) \ / \ (\mathbf{NIR} + \mathbf{SWIR})$$

3.1.4. Normalized differential wetness index (Ndwi)

NDWIs are generally sensitive to plant biomass and water conditions. NDWI is calculated from the green (G) and near-infrared (NIR) bands and emphasizes the quantity of water in a water body. High NDWI values indicate dense vegetation coverage and more vegetation water content. Low NDWI values mean low vegetation coverage and low vegetation water content. During water scarcity, the NDWI decreases. NIR wavelengths increase the high backscatter for vegetation and Soil and reduce the water's poor reflectivity. The result of the NDWI is positive for water and negative or zero for Soil and vegetation.

$$\mathbf{NDWI} = (\mathbf{G} - \mathbf{NIR}) / (\mathbf{G} + \mathbf{NIR}) \text{ Or } \mathbf{NDWI} = (\mathbf{NIR} - \mathbf{SWIR}) / (\mathbf{NIR} + \mathbf{SWIR})$$

3.1.5. Soil adjusted vegetation index (Savi)

SAVI is derived from the Red and NIR bands with a ground brightness correction factor (L), which has a value of 0.5 to account for all land cover varieties derived from the surface's reflectance.

$$\mathbf{SAVI} = ((\mathbf{NIR} - \mathbf{RED}) \ / \ (\mathbf{NIR} + \mathbf{RED} + \mathbf{L})) * (\mathbf{1} + \mathbf{L})$$

3.1.6. Lulc

Maps of land use and land cover (LULC) are essential for many different fields and industries. These maps play an important part in urban planning and development, helping with decisions regarding infrastructure development and zoning for land use.

They also play a significant part in environmental conservation by identifying and protecting natural habitats and ecosystems, keeping track of deforestation, and analyzing changes in vegetation cover. LULC maps improve crop planning and aid in soil suitability assessments in agriculture and land management, and they support watershed monitoring and flood risk assessment in water resource management. LULC maps are also employed in planning infrastructure and transportation, disaster risk reduction plans, forestry and resource management, climate change research, and biodiversity conservation initiatives. These give crucial information for making knowledgeable decisions, managing land sustainably, protecting the environment, and addressing various land use and sustainability issues. The LULC Map of Kanyakumari District shown in **Figure 4**.

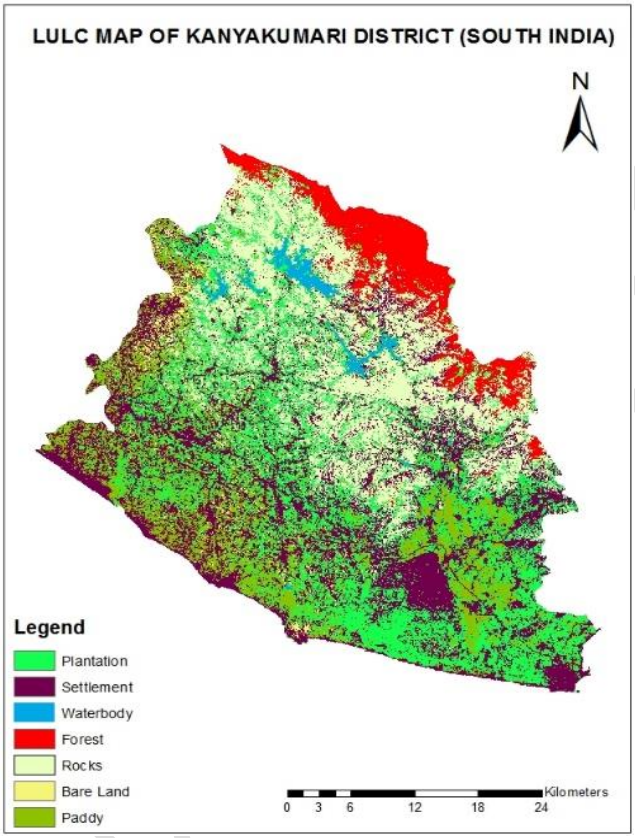


Figure 4. LULC Map of Kanyakumari District

3.1.7. Geology

The geology of the Kanyakumari district is distinguished by a wide range of rock formations that display a mixture of old crystalline rocks, coastal sediments, and laterite deposits. Specific geological hazards result from this region's geological makeup, closely related to tectonic activity and coastal influences. Laterite deposits, known for their iron-rich composition and valuable resource for construction, are found in the district's elevated regions. Sedimentary rocks, like sandstones and shales, are common along the district's coastline. Various geological processes shaped these rocks, including the dynamic forces of rivers and waves. Notably, the area is distinguished by charnockites, a variety of granitic rock distinguished by its distinctive reddish color and frequently used as building materials. The Geology Map of Kanyakumari District shown in **Figure 5**.

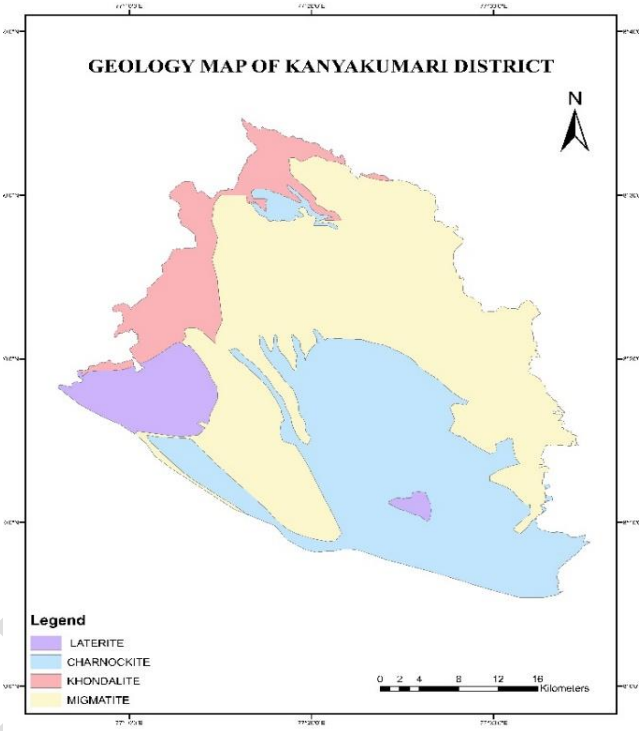


Figure 5. Geology Map of Kanyakumari District

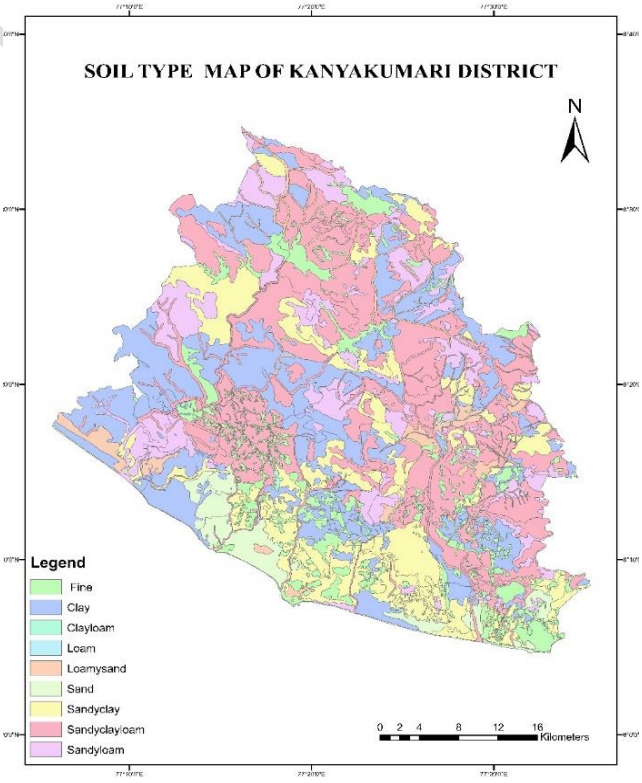


Figure 6. Soil Type Map of Kanyakumari District

3.1.8. Soil types

The diverse soil types in Kanyakumari district influence farming methods, crop patterns, and the overall ecological balance. Farmers and land managers must carefully plan

land use and conservation initiatives to maintain the region's natural resources and agricultural productivity. The Kanyakumari district has much red soil, especially in the hilly areas, distinguished by its reddish to brownish color due to its high iron oxide content. Although this soil type typically drains well, it is often nutrient-poor, making it suitable for growing crops like groundnuts, millets, and pulses with the correct irrigation and fertilization. In the hilly and forested areas of Kanyakumari district, lateral soils are commonly found on high plateaus. To improve the soil quality in these areas for agricultural use, farmers may need to add organic matter and nutrients. In the coastal plains of the Kanyakumari district, alluvial soils are prevalent along riverbanks and estuaries, formed by the deposition of clay and silt carried by rivers and streams. The Soil Type Map of Kanyakumari District shown in **Figure 6**.

Alluvial soils are fertile and ideal for growing rice, coconuts, and other cash crops because they retain moisture well. Along the Kanyakumari district's coastline, sandy soils are common, defined by their low water-holding capacity and coarse texture. These sandy soils are suitable for growing crops that can withstand salt, such as cashews, coconut, and certain vegetables. However, due to their low water retention, they may require regular irrigation. Some district areas have gravelly soil, especially on rocky terrain and steep slopes. Rock-filled lateritic gravelly soils are less conducive to agriculture. However, they are necessary to preserve the region's biodiversity and can support vegetation. Regur or cotton soil, also referred to as black soil, is relatively uncommon in the Kanyakumari district but found in some places.

Cotton, pulses, and oilseeds thrive in these soils to their high level of organic materials and nutrient levels. The district has gravelly soil, especially on rocky terrain and steep slopes. Rock-filled lateritic gravelly soils are less conducive to agriculture; however, they are necessary for preserving the region's biodiversity and supporting vegetation. Regur or cotton soil, also referred to as black soil, is relatively uncommon in the Kanyakumari district but found in some places. Cotton, pulses, and oilseeds thrive in these soils because of their high organic matter and nutrients.

4. Statistical analysis

The statistical values are analyzed using XLSTAT software for variables like SOC, NDVI, NDWI, NDMI, SAVI, and BSI. The relationship between one dependent variable (NDVI, NDWI, NDMI, SAVI, and BSI) and the self-supporting Random Forest regression model was used to compute the variable (SOC), and then applied to estimate the SOC Concentration. The Random Forest regression Model fits the residual plots and the R2 coefficient for SOC Variables.

Table 1. Relationship between Index and Satellite Data using Random Forest Regression Analysis

SL.NO	INDEX	LANDSAT 8	LANDSAT 9
1	NDVI	0.870	0.896
2	BSI	0.878	0.948
3	NDMI	0.878	0.896
4	NDWI	0.878	0.939
5	SAVI	0.881	0.896

Note: The table illustrates the Random Forest Regression (R^2) values between NDVI, BSI, NDMI, NDWI, and SAVI.

5. Results

5.1. Descriptive statistics between soc and other index

The statistics observed for NDVI, SAVI, BSI, NDMI, and NDWI are present in Table 1. The highest SOC is observed in the area of high vegetation, and the lowest SOC is obtained in the waterbodies of the study area. The relationship between the NDVI index, which represents the state of vegetation and land cover, and the SOC was evaluated using a Random Forest regression function, yielding highly reliable R^2= 0.870 for Landsat 8 and R^2= 0.896 for Landsat 9. The NDVI shows high values in the region of the northwest and low values in the Northeast part of the study area. In the NDVI Spectral range for Landsat 8, the Red Band is between 0.64 - 0.67 µm and for NIR Band is between 0.85 - 0.88 µm and for Landsat 9, the Red Band is between 0.636–0.673 µm and for NIR Band is between 0.851–0.879 µm. It has used hyperspectral datasets to compare differences in NDVI that directly influence the spectral band characteristic. It also demonstrates the importance of NDVI between Landsat geometries and vegetation management activity and growth throughout the year. The NDVI range for Landsat 8 is between -0.279566 and 0.609625, and for Landsat 9, it ranges from -0.166845 to 0.657373, illustrated in **Figure 7** a & b. Due to bandwidth, spatial resolution, and data processing differences NDVI, behavior can vary significantly across different sensors, especially between space-borne and airborne. When NDVI is high, the vegetation is high, so the SOC content in the Soil is high. The NDVI is high in the Southwest area in both Landsat 8 and 9, low in the east area and very low in waterbodies of the study area in Landsat 8 and 9.

The bare Soil is more if the BSI value is high. The Spectral range of BSI for Landsat 8 the Red Band is between 0.64 - 0.67 µm, the NIR Band is between 0.85 - 0.88 µm, the SWIR Band is between 0.450 - 0.51 µm and the Blue Band is between1.57 - 1.65 µm and for Landsat 9, the Red Band is between 0.636–0.673 µm and for NIR Band is between 0.851–0.879 µm SWIR Band is between 1.566–1.651 µm and Blue Band is between0.452–0.512 µm. The relationship between BSI, which indicates vegetation change, and SOC is R^2 = 0.878 for Landsat 8 and R^2 = 0.948 for Landsat 9. When BSI is high, the SOC is low in that area. The range of BSI for Landsat 8 ranges between -0.00253593 to 0.597988, and for Landsat 9 ranges from -

0.244995 to 0.272404, which is given in **Figure 8** a & b. When the NDVI is high, the BSI is very low; it shows high vegetation in the area, so the SOC is high. The Maximum BSI is obtained in the southeast region in both Landsat 8 and 9, and the minimum value is observed in the southwest area in Landsat 8 and North in Landsat 9 of the study area.

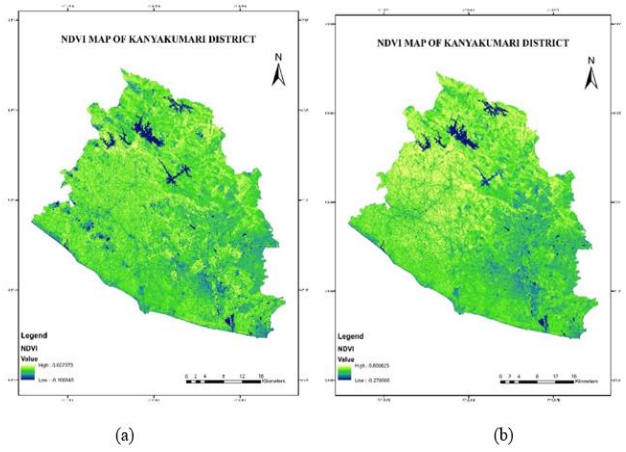


Figure 7a & b. NDVI Map of Kanyakumari District for Landsat 8 And 9

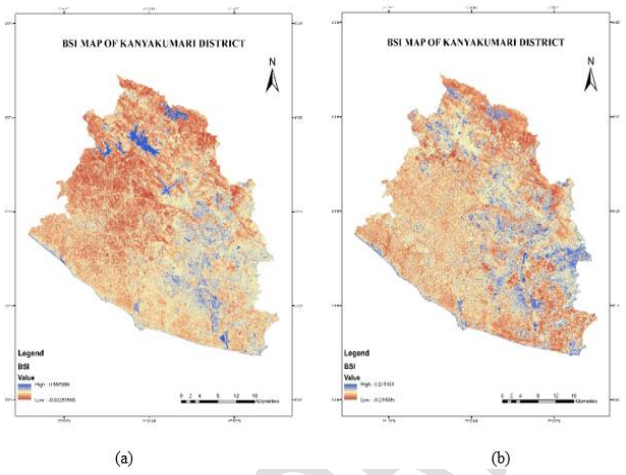


Figure 8 a & b. BSI Map of Kanyakumari District for Landsat 8 And 9

The relationship between NDMI, which indicates moisture content, and SOC is R^2 = 0.878 for Landsat 8 and R^2 = 0.896 for Landsat 9. The Spectral range of NDMI for Landsat 8 for NIR Band is from 0.85 to 0.88 µm The SWIR Band lies between 0.450 - 0.51 µm and for Landsat 9, for NIR Band is the SWIR Band, which is between 0.851 and 0.879 µm 1.566–1.651 µm. The NDMI for Landsat 8 is -0.322584 to 0.436183, and for Landsat 9, it is between -0.314451 to 0.375168, given in **Figure 9** a & b. When BSI is low, the moisture content in the area is high due to the presence of vegetation. So, when BSI increases, NDMI decreases, and vice versa. As the vegetation content is high in NDVI, when the NDVI increases, NDMI also increases. The Maximum NDMI is obtained in the Northwest area in the northeastern part of the research area is where the lowest value is seen in both Landsat 8 and Landsat 9, and North in Landsat 9.

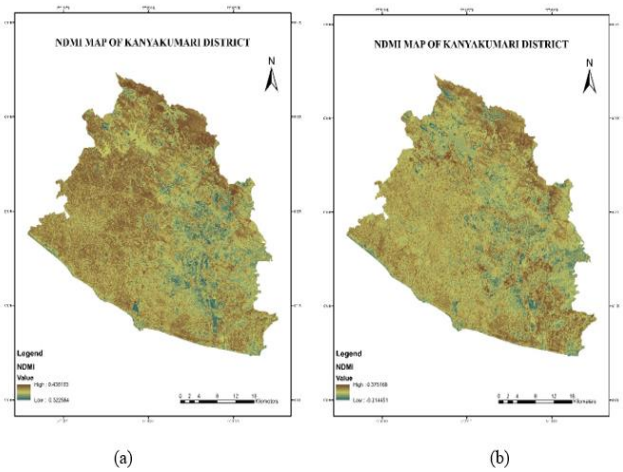


Figure 9 a & b. NDMI Map of Kanyakumari District for Landsat 8 And 9

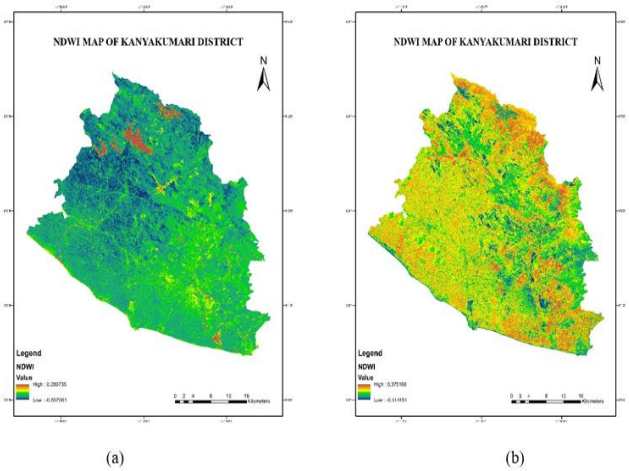


Figure 10 a & b. NDWI Map of Kanyakumari District for Landsat 8 And 9

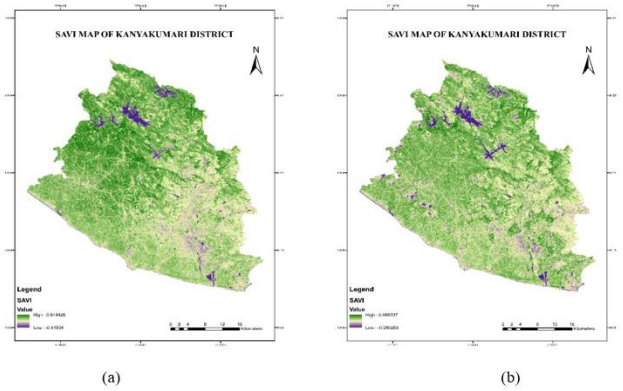


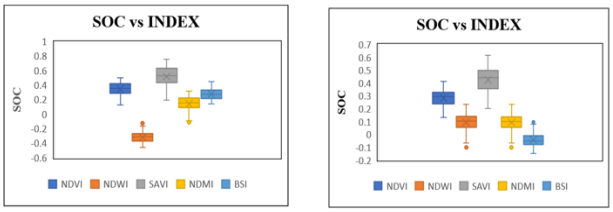
Figure 11 a & b. SAVI Map of Kanyakumari District for Landsat 8 And 9

The relationship between NDWI, which represents the humidity or wetness index, and SOC shows that Landsat 8 has R^2 = 0.878, and Landsat 9 has R^2 = 0.939. The NDWI Spectral range for Landsat 8, the Green Band is between 0.53 - 0.59 µm and for the NIR Band spans 0.85 to 0.88 µm and for Landsat 9, the Green Band is between 0.533–0.590 µm and for NIR Band is between 0.851–0.879 µm. The NDWI range for Landsat 8 is between -0.537361 and 0.280735 for Landsat 9 is between -0.314451 to 0.375168. The Maximum NDWI is obtained in the northern part of both Landsat 8 and 9, and the minimum value is observed

in the western part of Landsat 8 and the eastern part of the study area in Landsat 9. When NDWI values are greater than zero, they represent surfaces with water; when the NDWI is low or zero, they represent surfaces without water. Water levels are higher in bodies with more vegetation; hence, moisture levels are low. So, if NDVI rises, NDWI is low and vice versa, as indicated in **Figure 10** a & b.

The relationship between SAVI and SOC for Landsat 8 is R^2 = 0.861 and R^2 = 0.896 for Landsat 9. In the SAVI Spectral range for Landsat 8, the Red Band is between 0.64 - 0.67 µm and for NIR Band is between 0.85 - 0.88 µm and for Landsat 9, the Red Band is between 0.636–0.673 µm and for NIR Band is between 0.851–0.879 µm. The SAVI range for Landsat 8 is between -0.41934 to 0.914426, and for Landsat 9, it is between -0.250263 to 0.986037. When the NDMI is high, the SAVI is also high due to the vegetation present. The Maximum SAVI is obtained in the Northwest area in both Landsat 8 and 9, and the minimum value is noted in the research area's northeastern region in both Landsat 8 and 9, as illustrated in **Figure 11** a & b.

Kanyakumari district has a varied land use pattern. It includes agricultural land, forests, urban areas, coastal areas, and hilly terrain. A sizable amount of the district is devoted to agriculture, with the cultivation of crops like rice, coconut, rubber, and spices. The hilly areas, particularly in the Western Ghats, are covered with dense forests. The soil types in Kanyakumari district vary based on the topography. Alluvial soils are found in the coastal plains, making them suitable for rice cultivation. Red and laterite soils are prevalent in the hilly regions, which are not very fertile and require proper land management practices. Black soil is also found in some district areas, suitable for crops like cotton and oilseeds. The Western Ghats have rich forest soils that support diverse vegetation. The geological composition of the Kanyakumari district is influenced by its location at the Indian Peninsula's southernmost point. The Western Ghats, which run through the eastern part of the district, are primarily composed of Precambrian rocks like gneiss, granite, and schist. The coastal regions have sedimentary rocks, including sandstone and limestone. The presence of these different rock types contributes to the varied topography of the district, with rugged mountains in the east and a relatively flat coastal area in the west.



(a) (b)

Figure 12 a & b. Box Plotting of INDICES of Kanyakumari District for Landsat 8 And 9

5.2. Box plot for landsat 8 and landsat 9

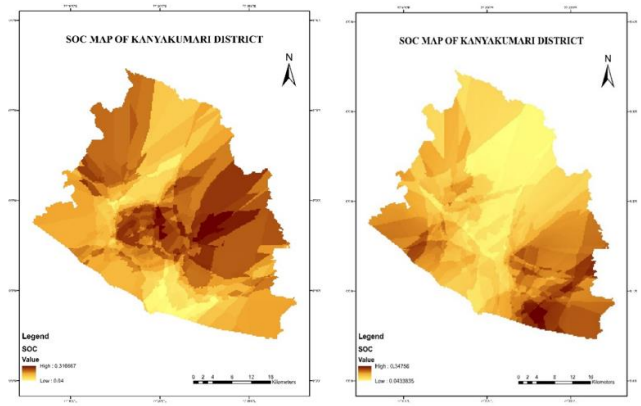
In **Figure 12** a & b, the box plot shows that the NDVI is higher in Landsat 8 than in Landsat 9, and SAVI is higher than NDVI as the soil brightness factor corrects it. The NDWI is higher in Landsat 9 than Landsat 8. The NDMI is nearly equal in both Landsat 8 and Landsat 9. The SAVI is higher in Landsat 8 than in Landsat 9. The BSI is high in Landsat 8 than Landsat 9.

5.3. Validation and prediction using machine learning model

A branch of artificial intelligence (AI), machine learning identifies patterns and connections in data to help guide decisions or actions. Computer programs that automatically find hidden patterns in data, predict outcomes, and improve performance based on prior experiences are known as machine learning algorithms. The Random Forest Regression Algorithm is the one that gives high accuracy and is used in our study.

Using a Random Forest model, we found a positive correlation between the SOC values, which are estimated and predicted, and the predicted SOC is mapped in **Figure. 13** a & b, and the results show that all the samples are randomly distributed. Using the RMSE and R^2 values, Predictive precision was estimated. The final approach is chosen according to which has a high RMSE and low R2 value. The digital SOC map helps to identify the SOC pixel-wise in the satellite imagery.

The predicted SOC Map of Landsat 8 shows that the SOC is very low in residential areas and high in forest areas. But in Landsat 9, the Forest areas have combined SOC from low to high, which shows that the areas with high vegetation have high SOC due to the undisturbed nature, and the areas with low SOC show that there is an influence of humans. The Landsat 8 has SOC ranges from 0.04 to 0.31, lower than Landsat 9. The Landsat 9 has SOC ranges from 0.0433835 to 0.34756, which is higher due to the penetration range or Spectral Resolution of Landsat 9.



(a) (b)

Figure 13 a & b. Predicted SOC Map of Kanyakumari District for Landsat 8 and 9

Agricultural lands, especially those under organic farming practices, have relatively higher SOC content due to incorporating organic matter into the soil. Crops like coconut, rubber, and spices may contribute to SOC accumulation. Forested areas, including the dense forests in the Western Ghats, often have substantial SOC content, as organic matter from plant litter and decaying vegetation enriches the soil. Coastal regions have lower

SOC content because of the sandy and well-draining soils, which are less conducive to organic matter retention. Alluvial soils found in coastal plains may have varying levels of SOC content, depending on land management practices and cropping systems. Well-managed paddy fields can accumulate organic carbon, while intensive cultivation may deplete it. Red and laterite soils in hilly areas generally have lower SOC content than other soil types due to their low organic matter content. Black soils have a relatively higher SOC content, especially in areas with sustainable agricultural practices and the incorporation of organic amendments. Forest soils, particularly in the Western Ghats, have the highest SOC content to the continuous input of vegetation-derived organic materials and little disturbance. The geological composition indirectly influences SOC content by affecting soil development and vegetation cover. Soils developed on different geological substrates may have varying organic carbon levels. In the Western Ghats, where Precambrian rocks predominate, the forested areas have higher SOC content due to the rich organic matter accumulation in these ecosystems. Coastal regions with sedimentary rocks have soils with relatively lower SOC content, primarily because of their sandy nature, which limits organic matter retention.

Table 2. Comparison of the existing models and proposed model

Authors	Methods	Accuracy
Bhunia, G. S., *et al*., (2019)	Multivariate Regression Model	89.63%
Lamichhane, S., *et al*. (2019)	Digital Soil Mapping Algorithm	90.50%
Kumar, U., *et al*., (2020)	Soil Physical-Chemical Quality Prediction	87.49%
Liu, Y. Y., *et al*., (2011)	Soil Moisture Prediction Using Microwave Data	91.22%
Proposed Model	RFR with Landsat Data	98.45%

Table 2 shows a comparison of existing models and the proposed model. The proposed technique maintains high accuracy levels of 98.45%. The proposed model enhances the total accuracy by 9.84%, 8.78%, 12.52%, and 7.92% better than existing techniques respectively. The comparison above indicates that the accuracy of the proposed model is superior to that of the existing models.

6. Discussion

The result indicated that Landsat 8 had lower SOC index correlation values than Landsat 9. In this research, 115 soil samples were verified for in-situ SOC, and the comparison of SOC indexes for Landsat 8 and 9 was carried out. In other research, SOC was usually estimated using field surveys and geostatistical modeling. The research mostly used in-situ observations. SOC is an important factor in the growth of plants, and it is predicted by correlations with indices like NDVI, BSI, NDMI, NDWI, and SAVI. Our findings authenticate significant correlations of SOC with NDVI, BSI, NDMI, NDWI, and SAVI. NDVI had a notably high correlation coefficient with SOC, which study that changes in NDVI have an important impact on SOC content. The spatial variability of BSI also contributes to SOC estimation, with the relationship being inverse; high NDVI is associated with low BSI and vice versa. It is interesting to observe that areas of high NDVI and SAVI in our study were also found to have higher SOC content, as depicted in the SOC maps. However, in our study, the SOC was notably linked with the NDVI index and SAVI Index, and the Regions with high NDVI are mapped to regions with high SOC and vice versa. For SOC Map preparation, the Random Forest model demonstrated an excellent model based on biomass generated. The relationship is identified between reference and predicted SOC, which are used practically. Landsat 8 offers 4,096 colors and 12-bit data, while 16,384 hues may be distinguished using Landsat 9. and provides 14-bit data of a given wavelength. Landsat 9's orbit is eight days out of phase, with Landsat 8 of 16 days to increase the temporal coverage of observations. Stray light correction, a problem found on Landsat 8's TIRS, A three-year design life Class-C instrument has been significantly improved by TIRS, a Class-B instrument with a five-year design life, is part of Landsat 9.

When there is an improved instrument, it gives good results. As a result, the predicted SOC is higher in Landsat 9 than in Landsat 8, which shows that the accuracy is higher in Landsat 9. By using SOC identification, the farmers can identify the level of SOC, and the plants are planted accordingly for high yield. A limitation of our study is that the Soil was sampled only once in the season. Therefore, for further studies, seasonal sampling is performed to determine SOC differences using Multispectral Remote Sensing and GIS (Geographic Information System).

7. Conclusion

This study used field data and Assessing the amount of organic carbon in top soil with remote sensing photos and Random Forest Regression Models (0 -20 cm). This study shows that factors such as NDVI, BSI, NDMI, NDWI, and SAVI help determine SOC. Comparing the large pixel size (30m) of Landsat 8 and 9 gives different results, in which Landsat 9 gave us comparatively high results. The NDVI and SAVI, the significant parameter, showed a good or nearly equal correlation with SOC, and the BSI and NDWI highly correlated with SOC. At the same time, the NDMI also had a good correlation with SOC. The parameters like soil type, LULC, and geology also affect the SOC in the study area, in which the soil type plays a significant role.

As a more modern satellite than its predecessors, Landsat 9 offers better spectral resolution, radiometric accuracy, and data quality. These improvements lead to more accurate red and NIR band measurements, which improves the accuracy of NDVI and SAVI calculations. Kanyakumari district's land use, land cover, soil types, and

geology are shaped by its unique geographical location and topographical diversity. This diversity supports various agricultural activities, natural resources, and regional tourism opportunities. This, thus, significantly affects the precision with which SOC content in crops may be anticipated using satellite data. So, creating a SOC Map helps us better determine where soil restoration and yield enhancement are needed. Kanyakumari district's SOC content varies widely across land uses, soil types, and geological areas. Forested regions and well-managed agricultural lands have higher SOC content, while coastal and hilly areas may have lower SOC levels. SOC deficiency has profound negative effects on soil fertility, structure, microbial activity, resilience to environmental stresses, and the global carbon cycle Effective land management practices can help enhance SOC content in the district, leading to improved soil fertility and sustainability in agriculture. For wise land management choices, conservation tactics, and attempts to reduce climate change, it is crucial to comprehend the relationship between SOC and the environment. Monitoring and maintaining healthy SOC levels contribute to sustainable agriculture, carbon sequestration, water resource management, and ecosystems' overall health and functioning. Managing and enhancing SOC levels using techniques like agroforestry, cover crops, organic amendments, and decreased tillage is crucial for preserving soil health and sustainability. Our ability to evaluate SOC levels and comprehend their environmental ramifications has drastically improved by using Landsat 9 and Landsat 8, and we are currently researching the possibility of combining data from several satellite missions to enhance SOC mapping and monitoring. Future SOC accuracy improvements necessitate the use of hyperspectral and microwave data. Furthermore, the paper recommends expanding the study to include field-level SOC prediction, which is beneficial for precision farming and helps improve environment management by providing food security.

References

Abdel-Kader, F. H. (2011), Digital soil mapping at pilot sites in the northwest coast of Egypt: A multinomial logistic regression approach, *The Egyptian Journal of Remote Sensing and Space Science*, **14**(1), 29–40.

Ahilan, A., Angel Sajani, J., Jasmine Gnana Malar, A. and Muthu Kumar, B. (2023), Machine Learning-Based Brain Disease Classification Using EEG and MEG Signals, In *International Conference on Frontiers of Intelligent Computing: Theory and Applications*, 487-498.

Alhameid, A., Ibrahim, M., Kumar, S., Sexton, P. and Schumacher, T. E. (2017), Soil organic carbon changes impacted by crop rotational diversity under no-till farming in South Dakota, USA, *Soil Science Society of America Journal*, **81**(4), 868–877.

Ashvanth Louison A. and Ben Sujitha, (2024), AMENET: Deep Learning-Based Automated Melanoma Detection Using Resnet50 and Random Forest, *International Journal of Current Bio-Medical Engineering*, **02**(01), 19–25.

Basset, C., Abou Najm, M., Ghezzehei, T., Hao, X. and Daccache, A. (2023), How does soil structure affect water infiltration? A meta-data systematic review, *Soil and Tillage Research*, **226**, 105577.

Bhunia, G. S., Kumar Shit, P. and Pourghasemi, H. R. (2019), Soil organic carbon mapping using remote sensing techniques and multivariate regression model, *Geocarto International*, **34**(2), 215–226.

Burrough, P. A., van Gaans, P. F. and Hootsmans, R. (1997), Continuous classification in soil survey: spatial correlation, confusion and boundaries, *Geoderma*, **77**(2–4), 115–135.

Cheng, L., Booker, F.L., Tu, C., Burkey, K.O., Zhou, L., Shew, H.D., Rufty, T.W. and Hu, S. (2012), Arbuscular mycorrhizal fungi increase organic carbon decomposition under elevated CO2, *Science*, **337**(6098), 1084–1087.

Clerici, N., Rubiano, K., Abd-Elrahman, A., Posada Hoestettler, J. M. and Escobedo, F. J. (2016), Estimating aboveground biomass and carbon stocks in periurban Andean secondary forests using very high resolution imagery, *Forests*, **7**(7), 138.

Coming, J. T. (2024), Environmental implications of modern food production: An analysis for the conscious consumer.

Franke, J. and Menz, G. (2007), Multi-temporal wheat disease detection by multi-spectral remote sensing, *Precision Agriculture*, **8**, 161–172.

Gilabert, M. A., González-Piqueras, J., Garcıa-Haro, F. J. and Meliá, J. (2002), A generalized soil-adjusted vegetation index, *Remote Sensing of environment*, **82**(2–3), 303–310.

Huang, S., Tang, L., Hupy, J. P., Wang, Y. and Shao, G. (2021), A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing, *Journal of Forestry Research*, **32**(1), 1–6.

Ismail, M. and Yacoub, R. K. (2012), Digital soil map using the capability of new technology in Sugar Beet area, Nubariya, Egypt, *The Egyptian Journal of Remote Sensing and Space Science*, **15**(2), 113–124.

Jaber, S. M., Lant, C. L. and Al-Qinna, M. I. (2011), Estimating spatial variations in soil organic carbon using satellite hyperspectral data and map algebra, *International Journal of Remote Sensing*, **32**(18), 5077–5103.

Jamalabad, M. S. (2004), Forest canopy density monitoring using satellite images. In *Geo-Imagery Bridging Continents XXth ISPRS Congress, Istanbul, Turkey, 2004*.

Kavitha, P., Ahilan, A., Gnanamalar, A. and Usha, M. (2025), Twin face recognition using a deep learning-based pixel difference network with edge maps, *Signal, Image and Video Processing*, **19**(1), 1–9.

Kumar, U., Mishra, V. N., Kumar, N., Srivastava, L. K. and Bajpai, R. K. (2020), Soil physical and chemical quality under long-term rice-based cropping system in hot humid eastern plateau of India, *Communications in Soil Science and Plant Analysis*, **51**(14), 1930–1945.

Lamichhane, S., Kumar, L. and Wilson, B. (2019), Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma*, **352**, 395–413.

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A., Wagner, W., Van Dijk, A. I. J. M., ... and Evans, J. P. (2011). Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, *Hydrology and Earth System Sciences*, **15**(2), 425–436.

Marchetti, A., Piccini, C., Francaviglia, R. and Mabit, L. (2012), Spatial distribution of soil organic matter using geostatistics:

A key indicator to assess soil degradation status in central Italy, *Pedosphere*, **22**(2), 230–242.

Masek, J. G., Wulder, M. A., Markham, B., McCorkel, J., Crawford, C. J., Storey, J. and Jenstrom, D. T. (2020), Landsat 9: Empowering open science and applications through continuity, *Remote Sensing of Environment*, **248**, 111968.

Mondal, S. and Chakraborty, D. (2022), Soil nitrogen status can be improved through no-tillage adoption particularly in the surface soil layer: A global meta-analysis, *Journal of Cleaner Production*, **366**, 132874.

Oldfield, E. E., Bradford, M. A. and Wood, S. A. (2019), Global meta-analysis of the relationship between soil organic matter and crop yields, *Soil*, **5**(1), 15–32.

Patel, H. V. (2018), *Mitigation of Recalcitrant Organic Matter from Municipal Solid Waste (MSW) Landfill Leachate by Chemically Enhanced Primary Treatment (CEPT)*, Lamar University-Beaumont.

Qiu, Y., Guo, L., Xu, X., Zhang, L., Zhang, K., Chen, M., Zhao, Y., Burkey, K.O., Shew, H.D., Zobel, R.W. and Zhang, Y. (2021), Warming and elevated ozone induce tradeoffs between fine roots and mycorrhizal fungi and stimulate organic carbon decomposition, *Science advances*, **7**(28), eabe9256.

Ruirui, C. and Xiaoting, W. (2022), BOOK REVIEW: Analytical methods for soil and agro-chemistry (in Chinese). Edited by HZ Zhu, PA He, CZ Chen, HM Zhou, DC Su, JM Xu, HY Qin, SD Bao, RK Lu SH Jiang Soil Science Society of China Beijing, China Agricultural Science and Technology Press, 2000, pp. 638. ISBN: 9787801199256, *European Journal of Soil Science*, **73**(4), e13280.

Somasundaram, J., Chaudhary, R. S., Awanish Kumar, D., Biswas, A. K., Sinha, N. K., Mohanty, M. ... and Chaudhari, S. K. (2018), Effect of contrasting tillage and cropping systems on soil aggregation, carbon pools and aggregate-associated carbon in rainfed Vertisols, *European Journal of Soil Science*, **69**(5), 879–891.

Sundarasekar, R. and Appathurai, A. (2022), Automatic brain tumor detection and classification based on IoT and machine learning techniques, *Fluctuation and Noise Letters*, **21**(03), 2250030.

Tisdall, J.M. and OADES, J.M. (1982). Organic matter and water-stable aggregates in soils, *Journal of soil science*, **33**(2), 141–163.

Vijayakumar, M. and Ahilan, A. (2024), An optimized chaotic S-box for real-time image encryption scheme based on 4-dimensional memristive hyperchaotic map, *Ain Shams Engineering Journal*, ***15***(4), 102620.

Villarino, S.H., Pinto, P., Jackson, R.B. and Piñeiro, G. (2021), Plant rhizodeposition: A key factor for soil organic matter formation in stable fractions, *Science Advances*, **7**(16), eabd3176.

Wright, S.F. and Upadhyaya, A. (1998), A survey of soils for aggregate stability and glomalin, a glycoprotein produced by hyphae of arbuscular mycorrhizal fungi, *Plant and soil*, **198**, 97–107.

Zeng, Y., Huang, W., Zhan, F., Zhang, H. and Liu, H. (2010), Study on the urban heat island effects and its relationship with surface biophysical characteristics using MODIS imageries, *Geo-spatial Information Science*, **13**(1), 1–7.