

Integrated Porosity and Pollutant Tracking in Engineering Geology: A Multi-Class Classification Supervised Machine Learning Approach - A Case Study in Ilorin, Kwara State, Nigeria

Dr Akinrinmade, A.O^{1*}, Smolic, H.², Olasehinde, D.A.³, Prof. Ige, O.O.⁴

¹Department of Geology and Mineral Sciences, Al-Hikmah University, Nigeria

²Graphite Note international, Ireland

³Department of Agricultural and Biosystems Engineering, Landmark University, Kwara State, Nigeria,

⁴Nigerian Geological Survey Agency, Utako District, Abuja, Nigeria

*Corresponding author

DOI: <https://doi.org/10.51584/IJRIAS.2024.907040>

Received: 25 June 2024; Revised: 18 July 2024; Accepted: 23 July 2024; Published: 13 August 2024

ABSTRACT

The world today faces unprecedented environmental challenges, including climate change, clean water scarcity, ocean contamination, and groundwater pollution, largely due to inadequate technology for tracking environmental pollutants. This study introduces a Supervised Machine Learning (SML) framework using multi-class classification to assess porosity and pollutant tracking in Ilorin, Kwara State, Nigeria. The study area is located within latitudes 8°44'6"N and 7°59'40"N and longitudes 4°09'40"E and 5°14'8"E, all within Nigeria's basement complex.

This research formulates a robust SML model for multi-class classification, categorizing different environmental suitability levels based on porosity, pollutant tracking, and environmental factors. The case study in Ilorin demonstrates the model's effectiveness, contributing significantly to the field of engineering geology.

A comprehensive approach integrates geological, geotechnical, geophysical, and environmental datasets. Surface and subsurface investigations, combined with supervised SML methods, predict suitability for porosity and pollutant tracking, providing insights into complex relationships impractical for manual analysis.

The study area includes Sokoto1, Sokoto2, Malete, Oke Oyi, Jimba, Omu Aran, and Ijagbo, all within Kwara State, which experiences cyclical dry and rainy seasons. Environmental factors considered include geological, geotechnical, geophysical, land use, water surface, and slope aspects.

The predictive model, utilizing multi-factorial analysis, categorizes outcomes into Highly Suitable, Moderately Suitable, and Not Suitable. Key factors influencing porosity and pollutant tracking include Suitability, Environmental Factors, Sub-Factors, Rating, Percentage of Influence, and Class. Model performance evaluation includes a fit analysis and confusion matrix.

The predictive model, trained on diverse environmental datasets, effectively categorizes suitability levels for porosity and pollutant tracking. The study identifies candidate sites with higher porosity and lower permeability, demonstrating practical applicability in decision-making processes for environmental analysis

and engineering geology related to surface and underground pollutant tracking.

Keywords: Supervised Machine Learning, Artificial Intelligence, Classification, Environmental, Porosity, Pollutant Tracking, Engineering Geology modelling, Nigeria.

INTRODUCTION

The profound interrelation between geological characteristics and environmental quality emphasizes the critical need for advanced technological integration in the field of engineering geology. This research work seeks to introduce a pioneering approach — a Supervised Machine Learning (SML) framework — for the integrated assessment of porosity and pollutant tracking within the geological landscape of Ilorin, Kwara State, Nigeria. Notably, Convolutional Neural Networks (CNN) have demonstrated success in pattern recognition for signals and images, with applications ranging from music genre classification (Li et al., 2010) to diverse everyday scenarios.

The central aim of this study is to formulate and execute a robust Supervised Machine Learning (SML) model tailored for multi-class classification. The primary focus is on effectively categorizing geological formations based on their suitability for porosity and pollutant tracking, with a simultaneous consideration of other relevant environmental factors. To illustrate the effectiveness of this approach, we conduct an in-depth case study in Ilorin, Kwara State, aiming to unveil nuanced insights into the intricate geological and environmental dynamics of the region.

The practical application of Artificial Intelligence (AI) has become widespread, extending beyond theoretical concepts to impact various fields. In contrast to human brains, which can discern between numerous items through deductive learning, AI algorithms require exposure to thousands of patterns before making appropriate decisions (Gurney, 2018). Addressing the challenges posed by the escalating human population's demand for natural resources, particularly in the face of geohazards, necessitates a comprehensive understanding of physical principles. AI's integration into geology, including mineral exploration, is underway, as evidenced by its applications in addressing challenges like associative memory, classification, pattern recognition, and optimization within geotechnical engineering using the Hopfield network (Saliu et al., 2020).

Machine learning algorithms, with their versatility, have significantly contributed to various earth science domains. In landslide assessments, (Marjanovic et al., 2011) showcase the impact of advanced algorithms in predicting, understanding, and mitigating landslide risks. Zhao and Wang (2020) demonstrate the advancement in soil property analysis through machine learning, aiding in enhanced land management and environmental planning. Geophysics benefits from machine learning in lithology classification based on geophysical data, as illustrated by (Bressan et al., 2020), facilitating geological mapping and exploration.

In the oil and gas industry, machine learning optimizes production processes, resource recovery, and operational efficiency (Attanasi et al., 2020; Gaurav, 2017; Mohaghegh, 2020; Snodgrass & Milkov, 2020). Multi-class classification, a pivotal concept, enhances the understanding of complex datasets, categorizing samples based on existing information (Farid et al., 2014).

Ilorin, Nigeria, presents a diverse geological and environmental landscape, demanding a sophisticated methodology to assess the impact of these factors on porosity and pollutant dispersion. Waste management practices, including collection, storage, treatment, handling, and disposal, pose inherent risks of environmental pollution. Uncontrolled groundwater, particularly in landfill sites, adds complexity due to the challenging regulation of leachate migration, potentially leading to groundwater contamination and broader environmental consequences. A holistic strategy is imperative, as highlighted by Akinrinmade et al. (2020), to mitigate pollution and ensure environmental sustainability in the intricate interplay between waste management and groundwater dynamics.

Porosity stands as a pivotal indicator, offering insights into the permeability and fluid flow dynamics within geological formations. Its significance lies in its role in evaluating critical aspects like groundwater movement, soil stability, and the potential migration of pollutants. The effective tracking of pollutants within the geological matrix holds paramount importance for environmental conservation, safeguarding human health, and fostering sustainable development.

Given the inherent challenges in quantifying groundwater on a large scale, the utilization of artificial intelligence (AI)-based algorithms emerges as a valuable solution to generate essential data and forecasts for effective groundwater management. Machine learning (ML) has demonstrated its effectiveness in producing maps for groundwater management, as evidenced by the work of Barzegar et al. (2018). These AI-driven algorithms prove particularly beneficial in tracking groundwater pollutants, addressing hazards such as nitrate contamination and other associated risks.

Recent advancements in machine learning methods have contributed to the enhancement of pollutant prediction, specifically for O₃ and NO₂. Notably, Support Vector Machines (SVM) have outperformed Neural Networks (NN) in predicting daily maximum O₃ concentrations, showcasing the sophistication and efficacy of these algorithms in environmental forecasting (Chelani, 2010). This supports the role of AI-based approaches in improving our understanding and management of groundwater quality, thereby contributing to more effective environmental stewardship and risk mitigation.

Despite recent strides in the field of artificial intelligence (AI) and its application in earthquake prediction, incorporating sophisticated deep learning (DL) into this realm remains a challenge. This challenge is attributed to the limited availability of features for training complex models, coupled with the fact that a majority of earthquake catalogues are stored in simplistic tabular formats. The constraints in data representation and feature richness pose obstacles to the effective application of deep learning in earthquake prediction, as highlighted by Mignan and Broccardo (2020).

Nevertheless, there is a positive trend in the development of more reliable and efficient seismic monitoring algorithms, thanks to the integration of DL approaches. The work of Mousavi et al. (2020) attests to the acceleration of advancements in seismic monitoring, facilitated by the adoption of deep learning techniques. Despite the challenges, the evolving landscape of AI-based earthquake prediction promises to overcome existing limitations and contribute to more accurate and timely seismic assessments.

The emergence of Supervised Machine Learning (SML) marks a significant paradigm shift in the evaluation of geological parameters. SML algorithms, trained on diverse datasets encompassing geological, geotechnical, geophysical, and environmental information, possess the capability to reveal intricate relationships and patterns that might elude traditional analyses. This research advocates for the innovative and efficient application of SML to integrate porosity and pollutant tracking on trained data, showcasing its potential in advancing the understanding of geological phenomena. Despite the widespread applications of machine learning algorithms in various earth science disciplines, there is a noticeable scarcity of studies on multi-class classification in the context of water exploration and groundwater pollutant tracking, as exemplified by the limited work conducted by Engle & Brunner (2019). This research gap underscores the need for further exploration and development of machine learning methodologies specifically tailored to address challenges in the realm of porosity and pollutant tracking.

STUDY AREA DESCRIPTION

The scope of this study encompasses the regions of Sokoto1, Sokoto2, Maletе, Oke Oyi, Jimba, Omu Aran and Ijagbo within Kwara state, Nigeria. These geographical areas are positioned between latitudes 8°44'6"N and 7°59'40"N and longitudes 4°09'40"E and 5°14'8"E. Geographically located in the southwestern part of Nigeria, these regions are in close proximity to the central part of the country. The climate in this area follows a cyclical pattern of dry and rainy seasons, with an annual rainfall ranging from 1270 mm to 1524

mm. The rainy season predominantly occurs from April to October, with peak rainfall observed in June/July and October. According to Michaelaschloegl (2023), the highest monthly temperature is recorded in March, reaching around 32°C, while the lowest temperature occurs in August, approximately at 25°C.

Noteworthy rivers in the region include Asa, Agba, Alalubosa, Okun, Osere, and Aluko, with some of these rivers draining into either the River Niger or River Asa, as documented by Oyegun (1985). The general elevation of the land in the western part varies from 273 m to 364 m above sea level.

Geology of Study Area

The study area is situated within Nigeria's Basement Complex, as depicted in Figure 1. Geological mapping has unveiled the presence of three primary rock types underlying the region: Granite Gneiss, Biotite Gneiss, and Migmatite. The detailed geology can be further classified into surface and subsurface geology. Surface geology comprises clay, lateritic soil, and the crustal top layer, exhibiting variations across different locations. Often, the lateritic soil dominates the surface, concealing much of the underlying geological features of the region.

To construct the geological map of the area, a synthesis of data from field mapping, literature reports, maps from the Nigeria Geological Survey Agency, and IKONOS imagery was employed. The compiled geology map underwent a sequence of processes, including scanning, processing, and digitization. A robust database was established and seamlessly integrated into the map, consolidating information on lithology, icons, and interpretations. Various lithologies were identified at the sampling sites, and a Geographic Information System (GIS)-based database was meticulously crafted, encapsulating icons, lithology details, and interpretations. These lithologies underwent classification and ranking based on their suitability as a landfill site, adhering to the standards delineated in Table 1. Subsequently, the lithology vector map underwent conversion into a raster map for in-depth analysis, as illustrated in the diagram in Figure 1.

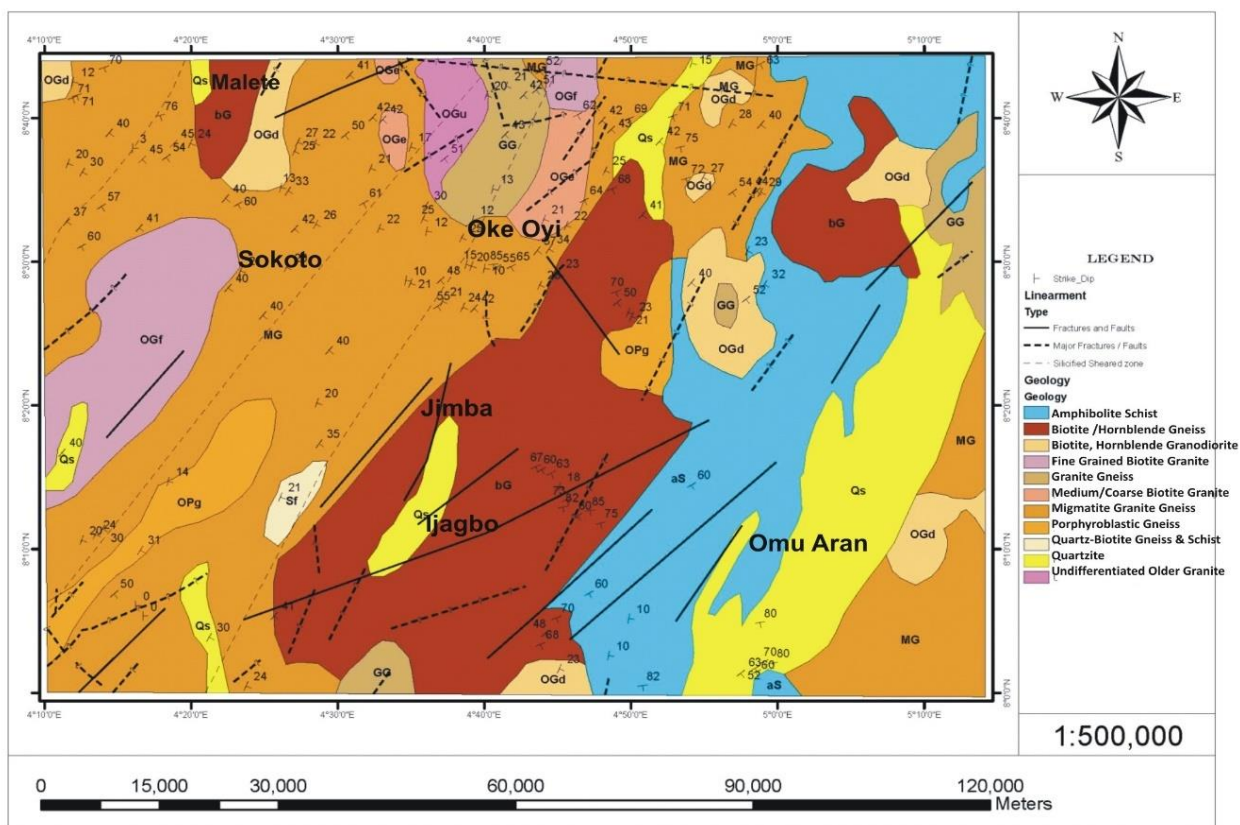


Figure 1: Geological Map of the study area

Geological Criteria

The geological characteristics of the area are intricately linked to the parent rock material, serving as the source of soil distribution. The suitability of the land for landfill and the movement of leachate hinge upon the strength and permeability of the soil. The structure of the rock body plays a pivotal role in influencing soil characteristics, parent rock permeability, and overall suitability, as illustrated in Table 1. The geological structure is also a critical factor in determining the movement of leachate and the potential for rock-slope failure, particularly along joints and inclined bedding.

To ensure the selection of appropriate parent rock materials and identify suitable areas for solid waste landfill, a thorough geological mapping process was undertaken. This mapping aimed to assess and understand the geological composition of the region, ensuring that the chosen parent rock materials meet the necessary criteria for effective and environmentally responsible solid waste disposal.

Table 1: Landfill suitability of bedrock (Oweis and Khera, 1998)

ROCK TYPE	SUITABILITY
Unfractured crystalline	Very high
Shale and clay	High
Limestone	Fair to poor
Sandstone	Poor to very poor
Unconsolidated sand/gravel	Unsuitable

The prevalent rock types in the area consist mainly of migmatite granite gneiss and biotite granite, exhibiting a coarse to medium-grained texture. In accordance with EPA guidelines from 2006, granite rock is deemed highly suitable for landfill, while the Migmatite-Gneiss complex is considered moderately suitable, and Quartzite is identified as the least suitable for landfill purposes. Notably, all the mapped sites satisfy the geological criteria necessary for the establishment of a sanitary landfill and other environmentally friendly sites, as detailed in Table 2.

Table 2: Rock Suitability Level

SITE	LOCATION	ROCK	SUITABILITY LEVEL	
S1&2	SOKOTO			
		1	Migmatite Granite Gneiss	Highly Suitable
S3	MALETE			
		1	Migmatite Granite Gneiss	Highly Suitable
S4	OKE OYI			
		1	Biotite Granite	Highly Suitable
		2	Granite Gneiss	Highly Suitable
	3	Porphyroblastic Gneiss	Moderately suitable	
S5	JIMBA			
		1	Granite Gneiss	Highly Suitable
	2	Biotite and Hornblend Gneiss	Moderately suitable	
S6	OMU ARAN			
		1	Migmatite Granite Gneiss	Highly Suitable
	2	Granite Gneiss	Highly Suitable	

	3	Quartzite	Least suitable
S7	IJAGBO	Biotite Gneiss	
		Biotite and Biotite Hornblende Gneiss	Moderately suitable

MATERIALS AND METHODS

The initial phase of the study involved a reconnaissance survey of selected sample locations, aiming to identify suitable areas for sanitary landfill site locations and areas for porosity and pollutants tracking in consultation with local relevant agencies. Subsequent steps included an environmental impact assessment, geological fieldwork, geophysical survey, and geotechnical soil sample collection for laboratory material testing and analysis. Site visits were conducted to ascertain soil thickness, lithology, and subsurface rock mass conditions.

The procedures for porosity and pollutant tracking for sanitary landfill site selection, as well as other environmental purposes and the availability of construction material, were divided into two categories: Surface and Sub-surface investigations. Surface investigations included desk studies, remote sensing for spatial data acquisition, and detailed geological mapping of the entire area to assess rock distribution, surface soil material, and environmental impact. Subsurface investigations involved geophysical and geotechnical data collection. The aeromagnetic and electrical resistivity techniques were employed for the study and delineation of geological subsurface structures. Vertical Electrical Sounding (VES) using the Schlumberger array was employed in geophysical investigations. For geotechnical data collection, soil samples were collected, labelled accordingly, and assessed for conditions based on British Standard International, 1377(1990). Twenty exploratory test pits were dug, and 60 soil samples were collected and analyzed.

In addition, a Supervised Machine Learning (SML) study on multi-class classification was utilized to predict the suitability of the area based on porosity and pollutant tracking, considering different environmental factors. Machine learning employed diverse methodologies to construct predictive models, excelling in tackling high-dimensional problems and categorizing rocks, soil, and environments. This methodology provided invaluable insights into complex relationships within datasets, which may be impractical or laborious to analyze through traditional, manual means.

These machine learning approaches were broadly categorized into two types: supervised and unsupervised learning methods. Supervised learning involved training models and making predictions based on rock types identified by geologists using labelled datasets. This method relied on labelled datasets where the algorithm learned from examples provided by experts, enabling it to make predictions or classifications when presented with new, unlabelled data. The utilization of supervised learning in geology facilitated the development of predictive models that aligned with expert knowledge, enhancing the understanding of rock classifications within the field.

Data Collection: Data Acquisition for GIS Database

The research utilized a diverse range of data and materials to comprehensively study the geological and environmental aspects of the region. The sources include IKONOS Imagery, Toposheets – RO1C07 to R18C13 (scaled at 1:50,000), LANDSAT ETM+ (2017, with a resolution of 28.5m, path and row wrs2- 190, 53) as outlined in Table 3. Additionally, ASTER imagery (with a resolution of 30m) covering the study area was employed to generate elevation and slope data. Geological information for the areas was gathered through geological fieldwork, satellite imagery, and data from the Nigeria Geological Surveys.

Geotechnical soil data were collected during site investigations and integrated with Soil maps to extract

details on soil types and their distribution. Topographical maps at a scale of 1:50,000 were utilized to outline the river systems within the area. IKONOS imagery played a crucial role in extracting information related to the built-up area, geological features, road networks, and validating water bodies in the study area. Furthermore, electrical resistivity surveys were conducted to delineate lithology and aquifer characteristics. Geometric data was collected through field surveys using a Global Positioning System (GPS Garmin-12). This comprehensive approach ensured a robust and multi-faceted dataset for the research.

Table 3: The Adopted Data and their Attributes

S/n	Data	Source	Year	Resolution	Relevance
1	Ikonos Imagery	Sat Imaging	2017	90m	2D Base Map
2	LANDSAT ETM	USGS	2016/2017	28.5m	Land use cover
3	ASTER DEM	Sat Imaging	2017	30.0m	3D Image (Terrain Analysis)
4	Geological map	NGSA	2017	Not applicable	Base Map
5	Aeromagnetic map	NGSA	2017	Not applicable	Inclination, Lineation fault
6	Topographical map	OSGF	2017	Not applicable	Base Maps
7	Drainage map	OSGF	2017	Not applicable	Drainage
8	Soil map	Field work	2017/2018	Not applicable	Soil distribution
9	GPS coordinates	Field	2017/2018	3m	Location Coordinates

Data Selection for Environmental Decision Factors

To construct the digital database for the sanitary landfill model in parts of South-West Nigeria, a diverse array of sources was employed. These sources encompassed geological, geotechnical, geophysical, environmental field data, and hydrological data of varying scales, as delineated in Table 4. The integration of data from these sources facilitated the development of a comprehensive and robust digital database, enhancing the accuracy and reliability of the sanitary landfill and pollutant tracking model for the specified region.

Table 4: Field and Spatial data used for porosity and pollutants tracking for sanitary landfill Modelling

S/N	Factors	Sub-factors	Sources	Information Used to create layers	Format	Scale or Resolution	Date
1	Geology	Distance to Faults Rock Exposure Porosity	Aeromagnetic Survey Ikonos Imagery/Field Mapping Geophysical Survey	Structures Geological/ Geotechnical Lithology	Digital Digital Digital	1:500,000	2017
2	Geotechnical/Soil	Type of Soil	Field work and Geotechnical Lab test	Type of Soil	Digital	1:500,000	2017

3	Geomorphology	Slope	ASTER Image	Elevation	Digital	1:500,000	2017
4	Water-Surface	Distance to Rivers	Hydrology Report: Ikonos Image	River, stream and Dams	Digital	1:500,000	2017
5	Water-Underground/ Geophysics	Distance to Wells Aquifer Flow Aquifer Vulnerability	Hydrology Report Field Geophysical survey	Aquifer Flow Classes Aquifer Vulnerability Classes	Digital Digital Digital	1:500,000	2018
6	Road	road	IKONOS	Road Network	Image	1:500,000	2017
7	Build-up Area	Build-up are	IKONOS	Settlement, land use and water body	Image	1:500,000	2017

Definition of Classes, Rating and Ranking

In this meticulous analytical study, the assignment of classes was executed with precision, considering the specific and pertinent conditions within the research area. The allocation process, detailed in Tables 5 and 6, carefully considered the unique contextual factors and parameters relevant to the research environment. This thorough approach to class assignment ensures a nuanced and contextually accurate representation of the studied phenomena, contributing to the robustness and reliability of the analysis within the specified research framework.

The eighteen sub-factors used in the Environmental Suitability Model (ESM) were categorized into classes for the Sanitary Landfill Environmentally Suitable Model. Ratings were placed on a scale of 1 to 10, as depicted in Table 7, where 1 represented the lowest level of suitability, and 10 indicated the maximum level of suitability for impact on the environment. The chosen measurement cycles of 1 to 10 were based on existing scales by Alavi et al. (2012), Hughes et al. (2005), and other relevant literature. However, it is noteworthy that the significance of each class may vary based on the location of interest and the unique features of the area, as highlighted by Al-Hanbali et al. (2011).

Table 5: Environmental Criteria for Buffer Zones rating interval

S/NO	CRITERIA (with respect to distance)	RECOMMENDATIONS (With References)
1	LAKE	≥ 60m (Nathanson, 2007) ≥ 300m (Bagchi, 1994; USEPA, 2005)
2	SLOPE	≤ 15° EPA., 2006, Flat area (Bagchi,1994; Montgomery, 2000; Gentle slope 10° -20° (Hughes <i>et al.</i> , 2005)
3	FLOWING STREAM	>90m (Bagchi, 1994); ≥150 (World Bank, 2004)
4	HIGHWAY	≥150m(Howard and Remson, 1978) ≥167m(WRSC, 1992); ≥500m(Zuquette <i>etal</i> ,2005)
5	WATER SUPPLY WELL	≥500(World Bank, 2004) ≥800m(bell,1999)
6	AIRPORT	≥330m(WRSC, 1992) ≥3048m(Bagchi,1994)
7	FLOODING FREQUENCY	

		100years (WRSC, 1992; Bagchi, 1994)
8	NEAREST SETTLEMENT	>500m(Bagchi 1994) >250m(World Bank, 2004) 1000m (Allen 2001)
9	DEPTH TO WATER TABLE (From the base of Mineral seal)	>0.6m(Howard and Ramson, 1978) 1.57(WRSC, 1992); ≥3.0(Frempong, 1999) >6.0m(Zuquette <i>et al</i> ,2001) >1.5m(Nathanson, 2000; World Bank, 2004)
10	DEPTH TO BASEMENT ROCK	3.3m (WRSC, 1992) >5m(Zuquette <i>et al</i> ,2005)
11	PROXIMITY TO FAULT	≥33m(WRSC, 1992); 60m (Nathanson, 2007)
12	PROXIMITY TO SINKHOLE	≥250m (WRSC, 1992)
13	PROXIMITY TO SOCIAL AMENITIES (POLES,GAS, WATER PIPES etc)	167m(WRSC, 1992; World Bank, 2004)
14	ACCESSIBILITY	30minutes drive or 10km from source (World Bank,2004)

Table 6 Suitability classes for different criteria under study (Sener 2005; EPA 2006; Leao *et al.* 2004)

Criteria	Class/ buffer zone	Suitability
Proximity to faults (m)	0-60	Very Low
	60-500	Low
	500-4000	Moderate
	4000-8000	High
	>8000	Very High
Proximity to roads (m)	0-100	Very Low
	100-700	Low
	700-1500	Moderate
	1500-4000	High
	4000-7000	Very High
Proximity to airports (m)	0-3000	Very Low
	3000-4000	Low
	4000-5000	Moderate
	5000-7000	High
	7000-30000	Very High
Proximity to Residential area (m)	0-3000	Very Low
	3000-5000	Low
	5000-6000	Moderate
	6000-8000	High
	>8000	Very High
Proximity to Wadies (m)	0-300	Very Low
	300-400	Low

	500-1000	Moderate
	1200-2000	High
	>2000	Very High
Proximity to Coast (m)	0-5000	Low
	5000-7000	Moderate
	>7000	High
Proximity to GW wells (m)	0-500	Very Low
	500-800	Low
	800-1200	Moderate
	1200-2000	High
	>2000	Very High
Proximity to Surface water (m)	0-500	Very Low
	500-1000	Moderate
	>1000	High
Permeability of strata	Coarse texture	Very Low
	Moderately coarse texture	Low
	Medium texture	Moderate
	Moderately fine texture	High
	Fine texture	Very High
Ground Water Depth (m)	0-10m	Very Low
	10-20m	Low
	20m-40m	Moderate
	40m-50m	High
	>50m	Very High
Geology	Dunite & schist	Very High
	Limestone & dolostone	Low
	Alluvial fans and terraces	Very Low

Table 7: Rating classes for sub-factors in the Modelling

Factors	Sub-factors	Class	Rating	% of Influence	
Geology	Distance to faults	<500 m	1	No Applicable	
		500 – 1000 m	2		
		1000 – 1500 m	3		
		1500 – 2000 m	4		
		2000- 2500 m	5		
		2500– 3000 m	6		
		3000– 3500 m	7		
		3500– 4000 m	8		
		4000– 4500 m	9		
		>4500 m	10		
		Porosity of Rock	Highly weathered rock		1
			Moderately weathered rock Fresh rock		5 10

Geotechnical/Soil	Type of Soil	No soil	1	20
		Peat	2	
		Gravel	3	
		Gravelly sand	4	
		Sand	5	
		Loamy sand	6	
		Sandy clay	7	
		Silty clay	8	
		Clay	10	
		Water Resources - Surface	Rivers and Stream	
500 – 1000 m	2			
1000 – 1500 m	3			
1500 – 2000 m	4			
2000- 2500 m	5			
2500– 3000 m	6			
3000– 3500 m	7			
3500– 4000 m	8			
4000– 4500 m	9			
>4500 m	10			
Water Resources - Underground/ Geophysical survey	Water Body: Lake, Dam and other man made water	<500 m	1	21
		500 – 1000 m	2	
		1000 – 1500 m	3	
		1500 – 2000 m	4	
		2000- 2500 m	5	
		2500– 3000 m	6	
		3000– 3500 m	7	
		3500– 4000 m	8	
		4000– 4500 m	9	
		>4500 m	10	
	Depth to rock	1m	1	20
		2m	2	
		3m	3	
		4m	4	
		5m	5	
		6m	6	
		7m	7	
		8m	8	
		9m	9	
		10m	10	
	Aquifer Vulnerability	High	2	
		Medium	6	
		Low	10	
Land use	Road/High way	<300m	4	3
		>300m	10	
	Built up Area	<300m	1	10
		500m	2	

		700m	3	
		900m	4	
		1000m	5	
		1200m	6	
		1400m	7	
		1600m	8	
		1800m	9	
		>2000m	10	
Slope	Elevation	<2°	10	4
		4°	9	
		6°	8	
		8°	7	
		10°	6	
		12°	5	
		14°	4	
		16°	3	
		18°	2	
		>20°	1	

Data Pre-processing

The identified problem is framed as a multi-class classification task, aiming to predict porosity and pollutant tracking based on a diverse set of environmental factors. The primary objective is to categorize or classify the levels of porosity and pollutant tracking into distinct classes or categories. The predictive modelling considers various environmental factors, spanning geological, geotechnical, geophysical, surface water, land use, and slope considerations.

The specific aspects of data pre-processing can be outlined as follows:

1. Environmental Factors

The dataset encompasses an extensive array of environmental factors, providing a thorough representation of various aspects. These factors include geological features, such as the proximity to fault lines, the presence of unconsolidated sand and gravel, the nature of sandstone, shale and clay porosity, and the characteristics of unfractured crystalline formations. Additionally, geotechnical information is incorporated, featuring the presence of different soil types like peat, gravel, gravely sand, sandstone, loamy sand, sandy clay, silty clay, and clay.

Geophysical parameters are considered in the dataset, including the proximity to boreholes and the depth to the rock strata. Surface water attributes, such as the distance to dams, lakes, and rivers, are included, alongside land use factors encompassing distances to residential areas, airports, roads, and farmlands. Moreover, slope-related information, specifically elevation, is incorporated into the dataset.

2. Sub-factor Classification

Each sub-factor within the environmental factors undergoes a meticulous classification process, segregating them into distinct classes based on their suitability measured in meters. This classification is executed by assigning a rating to each class on a percentage scale that spans from 1 to 10. A higher rating signifies an elevated level of importance for the respective sub-factor.

3. Suitability Levels:

The focal variable in the multi-class classification task is "POROSITY AND POLLUTANTS TRACKING." This variable is systematically categorized into distinct classes, namely Highly Suitable, Moderately Suitable, and Not Suitable.

5. Model Objective

The principal aim of the multi-class classification model is to achieve precise predictions of the class or category associated with porosity and pollutant tracking. This entails harnessing the collective influence of diverse environmental factors to construct a robust predictive model.

6. Target Variable Definition:

The target variable, "Porosity and Pollutants Tracking," is distinctly and precisely defined, demonstrating clarity in its formulation. The criteria for the multi-class classification are thoroughly established, outlining specific parameters and categories that serve to categorize and differentiate within the defined target variable.

7. Model Selection

To achieve precise predictions of "Porosity and Pollutants Tracking," a deliberate adoption of a supervised machine learning approach was implemented. The designated algorithm for this multi-class classification task is the Decision Trees classifier. This specific algorithm offers notable advantages, notably in furnishing interpretable insights into the intricate relationships among diverse environmental factors and the targeted outcome, as illustrated in Table 9a and 9b.

8. Data Splitting

In the creation of the Porosity and Pollutants Tracking model, a pivotal stage is the division of the dataset into training and testing sets. This step is vital to enable the model to grasp patterns from the training data and evaluate its effectiveness on unseen data. Employing a conventional 80-20 split, 80% of the data was designated for training purposes, while the remaining 20% was reserved for testing.

9. Model Training

The model underwent training with the training dataset to forecast suitability levels pertaining to porosity and pollutant tracking. Hyperparameter tuning was applied to enhance the model's performance, fine-tuning its parameters for optimal results. Cross-validation techniques were employed to fortify the model's robustness, ensuring its consistent and reliable performance across different subsets of the data as shown in Table 8a and 8b.

Table 8a: Model 80-20 Testing and Training Prediction

Train_Test_Type	Predicted Label	Predicted Probability	Predicted Correctness
test	Not Suitable	1	correct
test	Not Suitable	1	correct
test	Moderately Suitable	1	correct
test	Not Suitable	1	correct
test	Not Suitable	1	correct
test	Highly Suitable	1	correct

test	Highly Suitable	1	correct
test	Not Suitable	1	correct
test	Moderately Suitable	1	correct
test	Not Suitable	1	correct
test	Moderately Suitable	1	correct
test	Highly Suitable	1	correct
test	Highly Suitable	1	correct
test	Not Suitable	1	correct
test	Highly Suitable	1	correct
test	Moderately Suitable	1	correct
test	Not Suitable	1	correct
test	Moderately Suitable	1	correct
test	Not Suitable	1	correct

Table 8b: Model 80-20 Testing and Training Prediction

ENVIRONMENTAL FACTOR	SUB-FACTORS	CLASS (M)	RATING	% OF INFLUENCE	SUITABILITY	POROSITY AND POLLUTANTS TRACKING
LAND USE	Distance to Road	500	2	10	Very Low	Not Suitable
GEOTECHNICAL	No Soil	1	1	20	Very Low	Not Suitable
WATER SURFACE	Distance to Dam, Lake and Rivers	3000	6	22	Moderate	Moderately Suitable
SLOPE	Elevation	16	2	4	Very Low	Not Suitable
GEOTECHNICAL	Gravely Sand	60	4	20	Low	Not Suitable
LAND USE	Distance to Road	7000	9	10	Very High	Highly Suitable
GEOPHYSICAL	Proximity to Borehole	5500	10	21	Very High	Highly Suitable
WATER SURFACE	Distance to Dam, Lake and Rivers	1500	3	22	Low	Not Suitable
WATER SURFACE	Distance to Dam, Lake and Rivers	3500	7	22	Moderate	Moderately Suitable
LAND USE	Distance to Farm Land	1000	2	10	Very Low	Not Suitable
GEOPHYSICAL	Proximity to Borehole	3000	5	21	Moderate	Moderately Suitable
GEOLOGICAL	Unfractured Crystalline	2000	10	5	Very High	Highly Suitable

GEOPHYSICAL	Proximity to Borehole	5000	9	21	Very High	Highly Suitable
LAND USE	Distance to Residential Area	1000	3	10	Low	Not Suitable
GEOLOGICAL	Distance to Fault	6000	8	5	Very High	Highly Suitable
GEOPHYSICAL	Aquifer Depth to Rock	7	7	20	Moderate	Moderately Suitable
LAND USE	Distance to Farm Land	100	1	10	Very Low	Not Suitable
GEOTECHNICAL	Silty Clay	70	8	20	High	Moderately Suitable
LAND USE	Distance to Residential Area	100	1	10	Very Low	Not Suitable
WATER SURFACE	Distance to Dam, Lake and Rivers	1500	3	22	Low	Not Suitable

10. Model Evaluation

The effectiveness of the Porosity and Pollutants Tracking Model was rigorously assessed using a set of pertinent metrics tailored for multi-class classification. Key metrics included Accuracy, Precision, Recall, and F1-score for each class, accompanied by a detailed examination through a Confusion Matrix. The evaluation process exclusively leveraged the testing dataset, ensuring a comprehensive understanding of the model's generalization performance.

RESULT AND DISCUSSION

In the realm of Engineering Geological research, the development of an accurate predictive model for porosity and pollutant tracking is of paramount importance. This endeavour seeks to utilize a comprehensive dataset encompassing a myriad of environmental factors, including geological, geotechnical, geophysical, surface water, land use, and slope, to construct a robust model capable of providing insightful predictions.

The first set of factors considered falls under the geological category. These include the distance to fault, unconsolidated materials, and specific porosity-related attributes such as Sand and Gravel, Sandstone, Shale, and Clay Porosity. Additionally, Unfractured Crystalline formations are considered. Each sub-factor is meticulously assessed based on their suitability, rated in percentage from 1 to 10, with 1 being of least importance and 10 being of utmost significance. The resulting suitability levels are then categorized into very low, low, moderate, high, and very high classes.

Moving on to the geotechnical factors, considerations involve the presence or absence of specific soil types such as peat, gravel, gravely sand, sandstone, loamy sand, sandy clay, silty clay, and clay. Similarly, the suitability of these factors is determined and categorized into classes.

Geophysical factors, such as proximity to borehole and aquifer depth to rock, are also considered. Their suitability is assessed and classified accordingly.

Surface water factors, including Distance to Dam, Lake, and Rivers, are evaluated for their impact on porosity and pollutant tracking. The suitability of these factors is then rated and grouped into classes.

Land use factors, comprising Distance to Residential Area, distance to airport, distance to road, and distance to farm land, are examined for their influence on the outcomes. Suitability assessments are made, and classes are assigned accordingly.

Slope, measured in degrees, is the final factor under consideration. Its impact on the predictions is analysed, and suitability classes are assigned based on the degree of slope.

In the end, the predictive model aims to determine the suitability of the combined environmental factors for porosity and pollutant tracking. The outcomes are categorized into Highly Suitable, Moderately Suitable, and Not Suitable, providing valuable insights and actionable recommendations for informed decision-making in the field of Engineering Geological research.

1. SUITABILITY:

a. Feature name: SUITABILITY

b. Importance: Very important

c. Analysis and insights: Suitability is a critical factor significantly influencing the outcomes of porosity and pollutant tracking. A "Very High" suitability level enhances the likelihood of achieving the target outcome by a substantial multiplier of 34.0x.

d. Actionable insights to meet the goal: To attain the goal, prioritize environmental factors contributing to a "Very High" suitability level. Focus on creating conditions that align with the desired outcome, emphasizing the importance of suitability in the decision-making process.

2. ENVIRONMENTAL FACTOR:

a. Feature name: ENVIRONMENTAL FACTOR

b. Importance: Relatively important

c. Analysis and insights: Geological and geophysical factors positively impact outcomes, while water surface, slope, and land use have negative influences. Prioritizing positive factors through meticulous assessments can enhance porosity and pollutant tracking.

d. Actionable insights to meet the goal: Concentrate on geological and geophysical factors that have positive multipliers, addressing the negative impact of water surface, slope, and land use through targeted interventions. This involves a strategic focus on factors that contribute positively to achieve desired results.

3. SUB-FACTORS:

Analysis and insights: Sub-factors related to geology, geotechnical, geophysics, surface water, land use, and slope play a crucial role in influencing outcomes. Prioritizing favourable sub-factors, such as distance to fault and aquifer depth to rock, during site selection or project planning is essential.

4. RATING:

Higher ratings positively impact outcomes, underscoring the need to select sites or areas with favourable ratings.

5. Percentage (%) INFLUENCE:

Higher percentages of influence positively impact outcomes. Prioritizing environmental factors with higher percentages is crucial for optimizing porosity and pollutant tracking.

6. CLASS (M):

Sub-factor classification based on suitability in meters significantly impacts outcomes. Prioritizing higher classes is essential for achieving improved porosity and pollutant tracking.

Implementing these insights has enhanced the suitability of porosity and pollutant tracking, ultimately achieving the desired outcome of Highly Suitable. Further analysis of remaining features will provide a more comprehensive understanding for additional strategies.

Performance Analysis

In the evaluation of model performance, the identification of key drivers becomes pivotal, as it sheds light on the significance of each column or feature in making accurate predictions. The reliance of the model on specific columns, termed as key drivers, plays a crucial role in determining their importance. To quantify this importance, a permutation feature importance method was employed for calculation.

Among the various features considered, SUITABILITY emerged as the most crucial key driver in predicting POROSITY AND POLLUTANTS TRACKING. The parameters utilized during the model training encompass suitability level, environmental factors, sub-factors, rating, % of influence, and class. These parameters collectively contribute to the model's ability to accurately forecast outcomes related to porosity and pollutant tracking. Furthermore, it is observed that changes in SUITABILITY values directly impact POROSITY AND POLLUTANTS TRACKING outcomes. For instance, when the SUITABILITY is assessed as "very high," there is a corresponding increase in the likelihood of POROSITY AND POLLUTANTS TRACKING being classified as highly suitable by a specific numerical factor.

Suitability Level Impact Analysis

Understanding the impact of alterations in each individual feature is paramount, particularly in influencing the target feature – POROSITY AND POLLUTANTS TRACKING. Figure 2 provides a visual representation, offering insights into the current depiction of SUITABILITY's impact on POROSITY AND POLLUTANTS TRACKING.

The graphical representation utilizes a chart that vividly illustrates various suitability counts, namely highly suitable, moderately suitable, and not suitable, along the Y-axis. Simultaneously, it captures different suitability levels, including high, low, moderate, very high, and very low, along the X-axis. This visual depiction serves as a powerful tool for comprehending how the POROSITY AND POLLUTANTS TRACKING feature responds to diverse conditions of SUITABILITY.

By examining the chart, one can glean valuable insights into the intricate relationship between SUITABILITY and the POROSITY AND POLLUTANTS TRACKING feature. The varying levels of SUITABILITY are effectively mapped against different outcomes, providing a comprehensive understanding of how changes in SUITABILITY conditions influence the target feature. This analysis serves as a foundation for informed decision-making processes related to porosity and pollutant tracking, offering a nuanced perspective on the nuanced relationship between SUITABILITY levels and the ultimate outcomes of the model.

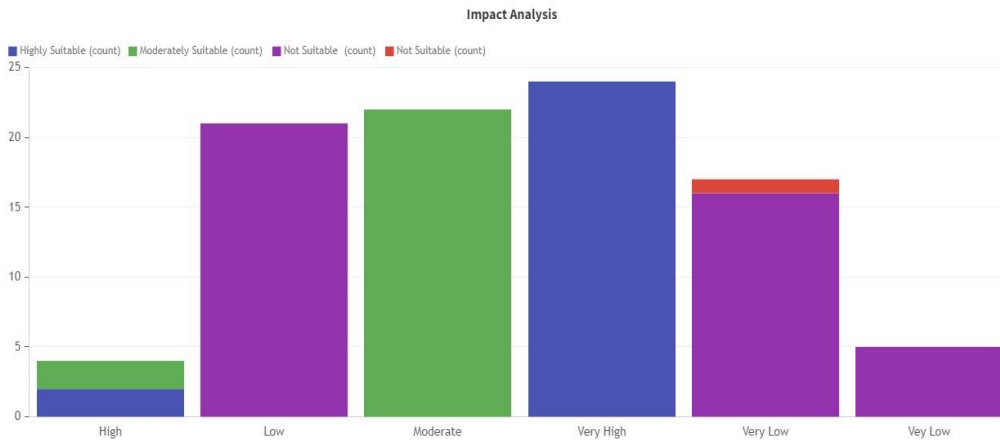


Figure 2: Suitability Level

Environmental Factor Impact Analysis

A thorough exploration of the model's analysis reveals a comprehensive understanding of how alterations in each environmental feature or factor, spanning geological, geophysical, geotechnical, land use, slope, and water surface, intricately influence the target feature – POROSITY AND POLLUTANTS TRACKING. The insights gleaned from this analysis serve to illuminate the nuanced relationships between various environmental elements and the ultimate outcomes of POROSITY AND POLLUTANTS TRACKING.

In Figure 3, a chart is presented to visually depict the impactful role of ENVIRONMENTAL FACTOR on POROSITY AND POLLUTANTS TRACKING. This chart serves as a graphical representation that vividly captures how variations in the ENVIRONMENTAL FACTOR, representing a collective amalgamation of geological, geophysical, geotechnical, land use, slope, and water surface factors, contribute to the fluctuations observed in the POROSITY AND POLLUTANTS TRACKING feature. The visual representation offered by this chart enhances the comprehension of the specific influences exerted by the ENVIRONMENTAL FACTOR on the target outcome.

This insightful visual analysis provides a foundation for understanding the intricate interplay between diverse environmental factors and their impact on the POROSITY AND POLLUTANTS TRACKING feature. It serves as a valuable tool for decision-makers, offering a clear understanding of the multifaceted influences that environmental factors can have on the predictive model. This clarity aids in the formulation of informed decisions and strategies aimed at optimizing POROSITY AND POLLUTANTS TRACKING based on the complex relationships identified through this comprehensive environmental factor impact analysis.

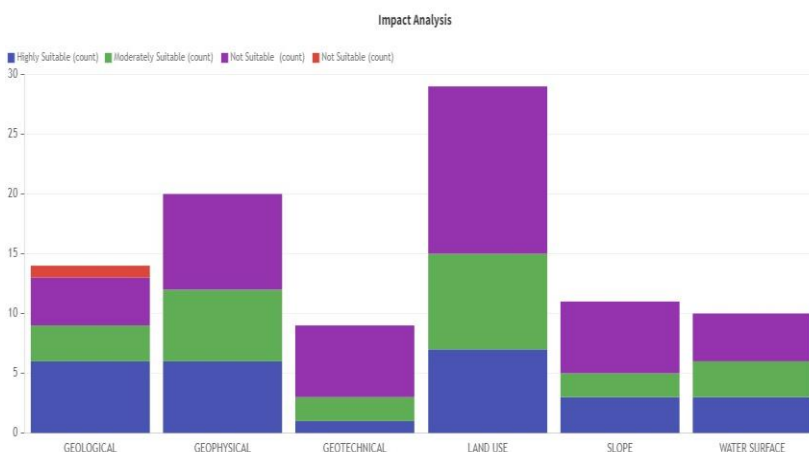


Figure 3: Environmental Factor

Sub-Factor Impact Analysis

The model analysis, as depicted in Figure 4, delves into a detailed examination of how alterations in each feature exert influence on the target feature, namely, POROSITY AND POLLUTANTS TRACKING. This analytical insight holds paramount importance for unravelling the intricate dynamics and dependencies between various features and the ultimate outcome of POROSITY AND POLLUTANTS TRACKING.

In Figure 4, a comprehensive chart is presented, visually representing the impactful role of SUB-FACTORS on POROSITY AND POLLUTANTS TRACKING. The SUB-FACTORS considered in this analysis encompass a diverse array, ranging from depth to rock, clay content, and distance to various landmarks (such as airport, dam, lake, rivers, farm land, fault, residential area, road) to elevation, gravel content, gravely sand, absence of soil, peat, proximity to borehole, sandstone, sandy clay, shale and clay porosity, silty clay, unconsolidated sand and gravel, and unfractured crystalline. This comprehensive chart vividly illustrates how variations in these SUB-FACTORS impact the POROSITY AND POLLUTANTS TRACKING feature.

Moreover, the chart goes a step further by categorizing the suitability level of these SUB-FACTORS, differentiating between highly suitable, moderately suitable, and not suitable. This classification provides a nuanced understanding of the intricate relationship between each SUB-FACTOR and its suitability in influencing the target outcome. Analyzing this visual representation becomes instrumental in making informed decisions and devising strategies to optimize POROSITY AND POLLUTANTS TRACKING. The specific characteristics of these SUB-FACTORS, as unveiled by the chart, serve as a valuable guide for refining strategies and interventions to achieve optimal outcomes in the realm of porosity and pollutant tracking.

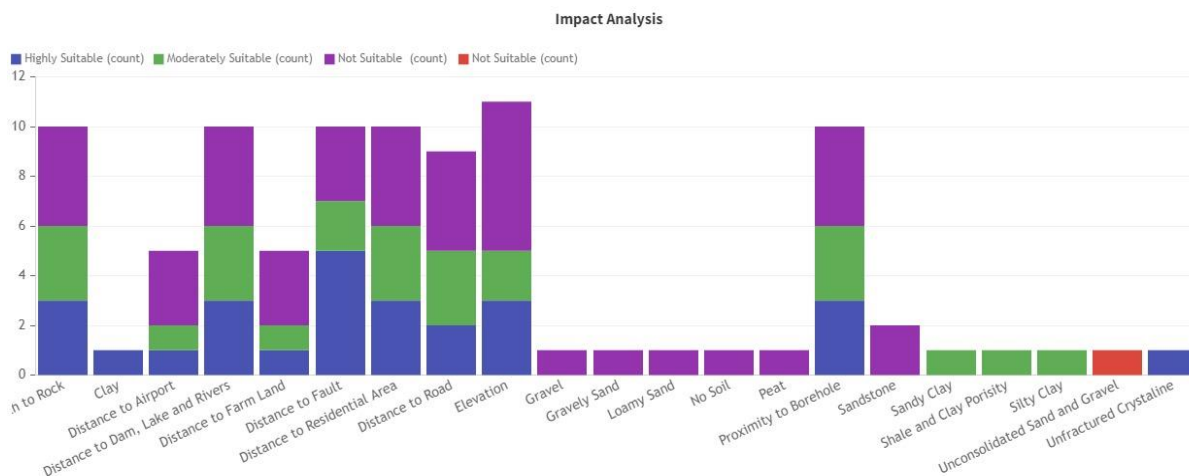


Figure 4: Sub-Factor

Rating Impact Analysis

The model analysis depicted in Figure 5 offers a comprehensive examination of the dynamic interplay between individual features and the target feature, POROSITY AND POLLUTANTS TRACKING. This analytical representation is instrumental in unravelling the intricate relationships and dependencies that influence the final outcome of POROSITY AND POLLUTANTS TRACKING.

Within Figure 5, the chart intricately illustrates the impact of the feature "RATING" on POROSITY AND POLLUTANTS TRACKING. The feature "RATING" is assessed on a scale ranging from 1 to 10, with 1 denoting the lowest level of suitability and 10 signifying the utmost level of suitability for its impact on the environment. This scaling system provides a nuanced perspective, allowing one to discern the varying degrees of influence that different ratings exert on the POROSITY AND POLLUTANTS TRACKING

outcome.

The chart visually categorizes the different rating levels, creating a clear representation of how each level contributes to the overall suitability and impact on the environment. One can use this information to make informed decisions about prioritizing areas with higher ratings and strategically selecting sites or regions that align with the desired level of suitability for POROSITY AND POLLUTANTS TRACKING.

This analytical insight aids in formulating targeted strategies and interventions to optimize POROSITY AND POLLUTANTS TRACKING based on the specific rating assigned to environmental conditions. The visual clarity provided by Figure 5 enhances the understanding of the role of the "RATING" feature in influencing the ultimate environmental outcome.

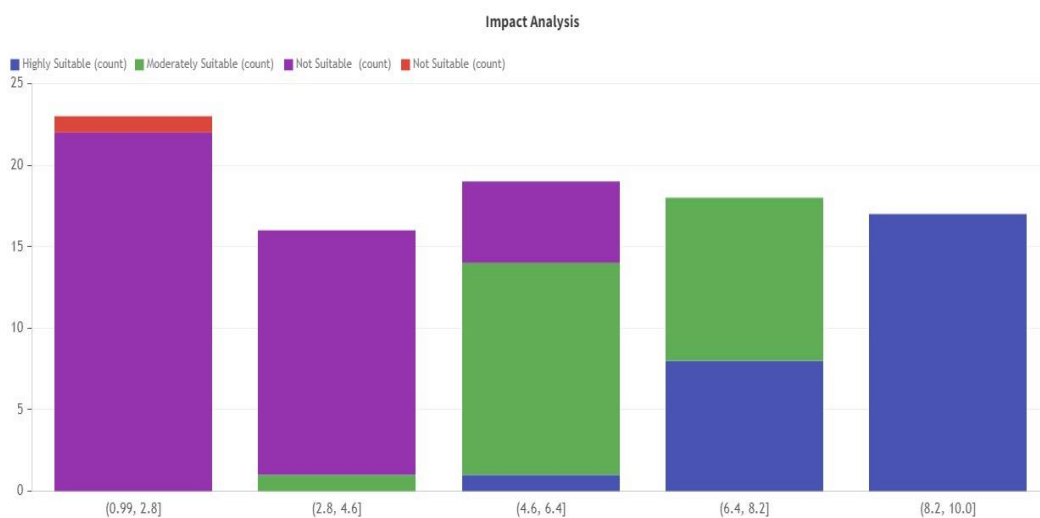


Figure 5: Rating

Percentage of Influence Impact Analysis

The model analysis presented in Figure 6 delves into the intricate dynamics of how alterations in individual features contribute to the overall impact on the target feature, POROSITY AND POLLUTANTS TRACKING. This detailed examination sheds light on the nuanced relationships and influences that shape the final outcome of POROSITY AND POLLUTANTS TRACKING.

Within Figure 6, the chart meticulously portrays the influence of the "PERCENTAGE OF INFLUENCE" feature on POROSITY AND POLLUTANTS TRACKING. The chart visually represents varying percentages of influence, offering a comprehensive understanding of how each percentage range affects the ultimate environmental outcome. The percentages are categorized into different bands, enabling one to discern the magnitude of influence exerted by each range on POROSITY AND POLLUTANTS TRACKING.

This graphical representation serves as a valuable tool for decision-makers, providing a clear depiction of the factors that wield the most significant influence on the environmental outcome. One can utilize this information to prioritize environmental features with higher percentages of influence, thereby optimizing strategies to enhance POROSITY AND POLLUTANTS TRACKING.

By focusing on the chart in Figure 6, stakeholders can formulate targeted interventions and actions to mitigate any negative impact associated with higher negative percentages of influence. The insights gained from this analysis empower stakeholders in making informed decisions to improve the overall suitability of POROSITY AND POLLUTANTS TRACKING based on the intricate interplay of the "PERCENTAGE OF

INFLUENCE" feature.

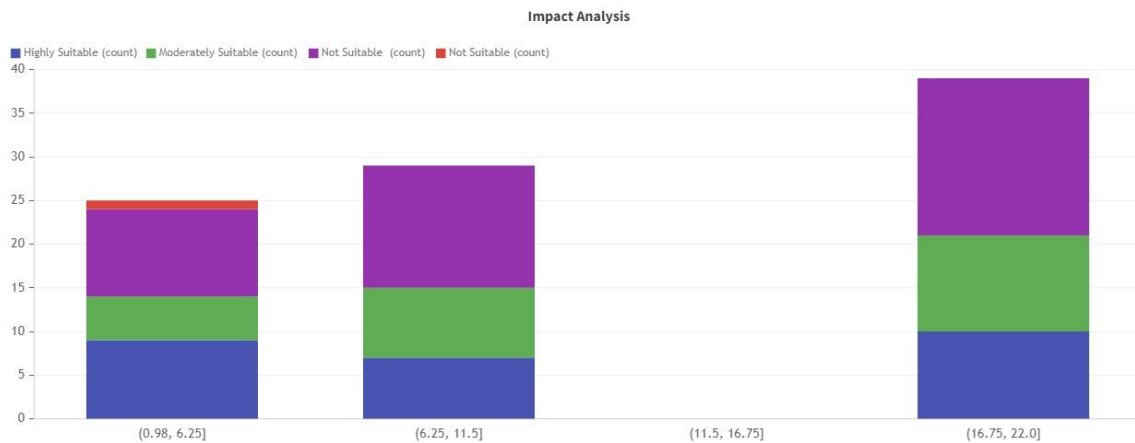


Figure 6: Percentage of Influence

Class Impact Analysis

The model analysis thoroughly examines the impact of alterations in each feature on the target feature, POROSITY AND POLLUTANTS TRACKING. Figure 7 visually represents this analysis, specifically focusing on the influence of the "CLASS (M)" feature.

In Figure 7, the chart elucidates the relationship between the "CLASS (M)" feature, displayed on the X-axis, and its impact on POROSITY AND POLLUTANTS TRACKING, illustrated on the Y-axis with various suitability levels. This visual representation allows one to discern how different classes within the "CLASS (M)" feature contribute to the overall environmental outcome.

By examining this chart, one can make informed decisions regarding the selection and prioritization of specific classes that align with the desired suitability levels for POROSITY AND POLLUTANTS TRACKING. This analysis provides a valuable tool for optimizing strategies and interventions, ultimately enhancing the overall environmental outcome based on the nuanced dynamics of the "CLASS (M)" feature.

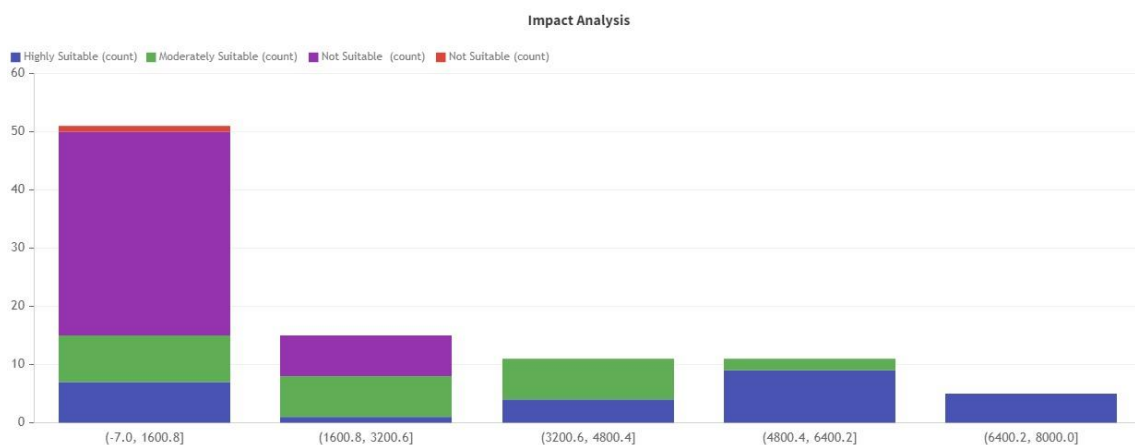


Figure 7: Class

The Model Fit depicted in Figure 8 gauges the performance of the model, revealing its effectiveness. By scrutinizing the outcomes for 19 rows in the test dataset, the comparison between the Model's accurate and inaccurate predictions for the POROSITY AND POLLUTANTS TRACKING column is illustrated.

A superior model fit is indicated by a higher percentage of correct predictions. The effectiveness of the

model is quantified by the accuracy of its predictions, emphasizing the importance of a higher correct percentage as an indicator of a well-performing model. Stakeholders can rely on this evaluation in Figure 8 to assess and validate the model's precision in predicting POROSITY AND POLLUTANTS TRACKING outcomes, contributing to informed decision-making in the context of environmental analysis.

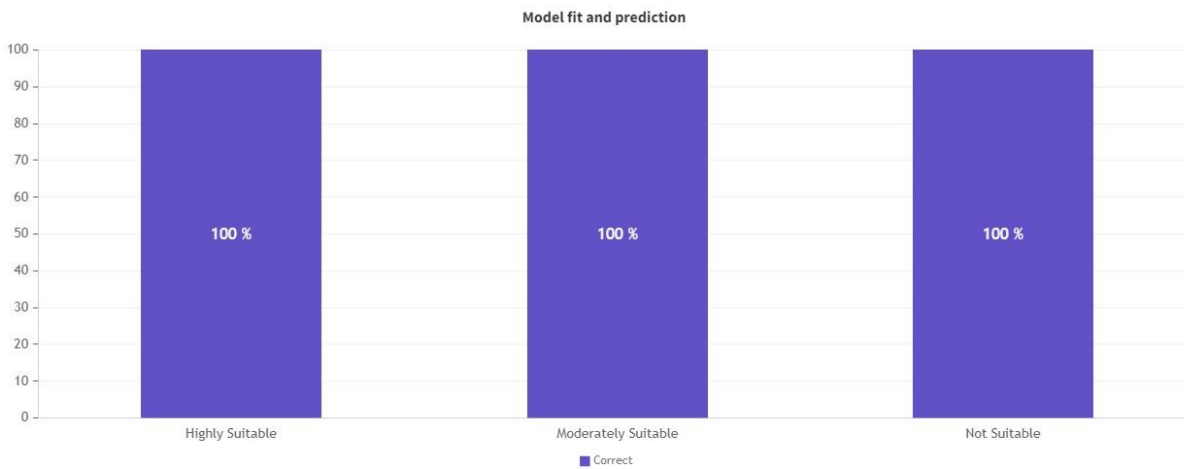


Figure 8: Model Fit

Confusion Matrix Analysis

The Confusion Matrix, depicted in Figure 9, serves as a powerful tool for unveiling classification errors within the model. It offers a clear visualization of whether the Model is encountering challenges in distinguishing between classes. For each class, the Confusion Matrix succinctly presents the number of correct and incorrect predictions. In this analysis, the Model predicted the column POROSITY AND POLLUTANTS TRACKING for a test dataset comprising 19 rows, and the predicted outcomes were compared to the historical outcomes.

The Confusion Matrix and corresponding model metrics collectively paint a picture of a highly accurate and precise model. The absence of incorrect predictions and the perfect F1 score signify a strong foundation, showcasing the reliability and effectiveness of the model in predicting POROSITY AND POLLUTANTS TRACKING.

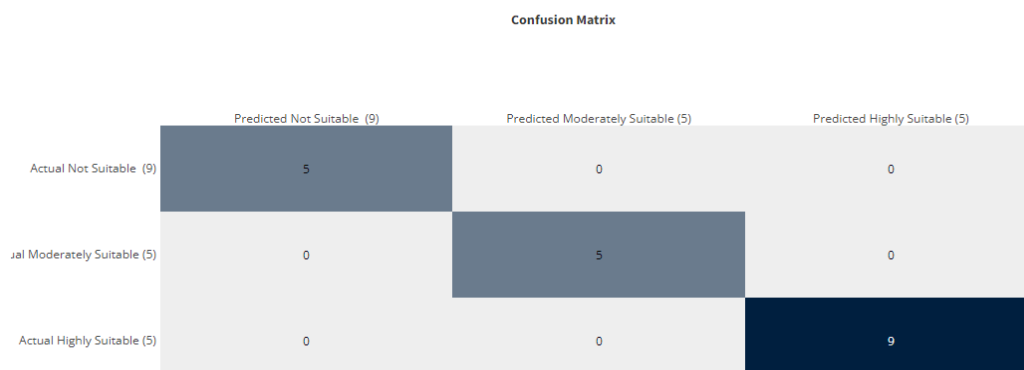


Figure 9: Model Accuracy Overview

Prediction with Trained Model

Having undergone successful training and deployment, the model is poised to execute predictions on specific datasets. To harness the predictive capabilities of this Model, a dataset must be chosen, adhering to the

following column requirements: ENVIRONMENTAL FACTOR, SUB-FACTORS, CLASS (M), RATING, % OF INFLUENCE, and SUITABILITY. Once the dataset is selected, the model will be applied to it, ushering in a transformative process where new columns will be appended to the dataset. These new columns will house the model's predictions and corresponding scores for each row.

This predictive phase is a culmination of the model's learning and training, now actively contributing to decision-making processes. The selected dataset's features, spanning environmental factors, sub-factors, class, rating, percentage of influence, and suitability, will be meticulously assessed by the model. Following this evaluation, the model will generate predictions, offering insights and predictions tailored to each row within the dataset as shown in Table 9 and 10. The table contains data with predicted probability.

The appended columns, showcasing the model's predictions and scores, will serve as valuable additions to the existing dataset. These additions not only provide the predicted outcomes but also offer a quantitative measure of the model's confidence or certainty in its predictions.

In essence, the prediction phase marks the practical application of the trained model's knowledge, translating it into actionable insights based on the specific dataset's characteristics as shown in Figure 10. This process empowers users to leverage the model's capabilities for informed decision-making, enhancing the overall utility and effectiveness of the predictive model in addressing queries related to environmental factors, suitability, and the intricate relationships within the dataset.

Model Multi-Class Classification on Envr Suitability Model Machine Learning

⊙ Predict ↶ API

Overall Model Performance predicting POROSITY AND POLLUTANTS TRACKING : ★ AM

* final model scores based on a test dataset (19 rows)

Decision Tree Classifier ★ 100.0% ★ 100.0% ★ 100.0%

Best Model ⊙ F1 ⊙ Accuracy ⊙ AUC ⊙

[Predict](#) [Predict Dataset](#)

The Model has been successfully trained and deployed and is now ready to make predictive scenario, please enter the values for the relevant factors in the 'what-if scenario' section below

What-if observation predicting POROSITY AND POLLUTANTS TRACKING

SUITABILITY
High

ENVIRONMENTAL FACTOR
GEOLOGICAL

SUB-FACTORS
Aquifer Depth to Rock

CLASS (M)
1

RATING
1

% OF INFLUENCE
1

[Predict](#)

- Aquifer Depth to Rock
- Clay
- Distance to Airport
- Distance to Dam, Lake and Rivers
- Distance to Farm Land
- Distance to Fault
- Distance to Residential Area
- Distance to Road
- Elevation

Figure 10: Multi—Class Classification Model Interface

Table 9: Multi-Class Classification on Environmental Suitability Model Machine Learning predicted

Predicted Label	Predicted Probability	% OF INFLUENCE	CLASS (M)	ENVIRONMENTAL FACTOR	POROSITY AND POLLUTANTS TRACKING	Predicted Correctness	RATING	SUB-FACTORS	SUITABILITY	Train test type
Not Suitable	100%	10	500	LAND USE	Not Suitable	correct	2	Distance to Road	Very Low	test
Not Suitable	100%	20	1	GEOTECHNICAL	Not Suitable	correct	1	No Soil	Very Low	test
Moderately Suitable	100%	22	3000	WATER SURFACE	Moderately Suitable	correct	6	Distance to Dam, Lake and Rivers	Moderate	test
Not Suitable	100%	4	16	SLOPE	Not Suitable	correct	2	Elevation	Very Low	test
Not Suitable	100%	20	60	GEOTECHNICAL	Not Suitable	correct	4	Gravelly Sand	Low	test
Highly Suitable	100%	10	7000	LAND USE	Highly Suitable	correct	9	Distance to Road	Very High	test
Highly Suitable	100%	21	5500	GEOPHYSICAL	Highly Suitable	correct	10	Proximity to Borehole	Very High	test
Not Suitable	100%	10	500	LAND USE	Not Suitable	correct	2	Distance to Road	Very Low	test
Not Suitable	100%	22	1500	WATER SURFACE	Not Suitable	correct	3	Distance to Dam, Lake and Rivers	Low	test
Moderately Suitable	100%	22	3500	WATER SURFACE	Moderately Suitable	correct	7	Distance to Dam, Lake and Rivers	Moderate	test
Not Suitable	100%	10	1000	LAND USE	Not Suitable	correct	2	Distance to Farm Land	Very Low	test
Moderately Suitable	100%	21	3000	GEOPHYSICAL	Moderately Suitable	correct	5	Proximity to Borehole	Moderate	test
Highly Suitable	100%	5	2000	GEOLOGICAL	Highly Suitable	correct	10	Unfractured Crystalline	Very High	test
Highly Suitable	100%	21	5000	GEOPHYSICAL	Highly Suitable	correct	9	Proximity to Borehole	Very High	test
Not Suitable	100%	10	1000	LAND USE	Not Suitable	correct	3	Distance to Residential Area	Low	test
Highly Suitable	100%	5	6000	GEOLOGICAL	Highly Suitable	correct	8	Distance to Fault	Very High	test
Moderately Suitable	100%	20	7	GEOPHYSICAL	Moderately Suitable	correct	7	Aquifer Depth to Rock	Moderate	test
Not Suitable	100%	10	100	LAND USE	Not Suitable	correct	1	Distance to Farm Land	Very Low	test
Moderately Suitable	100%	20	70	GEOTECHNICAL	Moderately Suitable	correct	8	Silty Clay	High	test
Not Suitable	100%	10	100	LAND USE	Not Suitable	correct	1	Distance to Residential Area	Very Low	test

Table 10: Suitability Level Summary of the Study Area

S/N	Site Name	Site No	Suitability Level
1	Sokoto1	S1	Not Suitable
2	Sokoto2	S2	Moderately suitable
3	Malete	S3	Moderately suitable
4	Oke Oyi	S 4	Not suitable
5	Jimba	S 5	Moderately suitable
6	Ijagbo	S 6	Not suitable
7	Omu Aran	S 7	Moderately Suitable
8	Outside Research Scope	S 8	Suitable
9	Outside Research Scope	S9	Suitable
10	Outside Research Scope	S10	Suitable
11	Outside Research Scope	S11	Suitable
12	Outside Research Scope	S11	Suitable

Upon completion of the predictive analysis, the model result in the study area, show sites that were classified as the most suitable landfill among several sites outside the scope of this research having met all the environmental criteria, four candidate sites were identified as moderately suitable landfill sites (Sokoto2, Malete, Jimba and Omuaran) among several sites within the scope of this research, while three candidate sites were identified as not suitable landfill sites (Sokoto1, Ijagbo, and Oke Oyi). The trained model was used in real world scenario and can be deployed for similar investigations.

CONCLUSION

The study presents a comprehensive and innovative approach to porosity and pollutant tracking in engineering geology, utilizing a multi-class classification supervised machine learning model. The research focused on the regions of Sokoto1, Sokoto2, Malete, Oke Oyi, Jimba, Omu Aran and Ijagbo within Kwara State, Nigeria, using an extensive dataset covering geological, geotechnical, geophysical, surface water, land use, and slope factors.

The importance of the suitability factor emerged as a key driver, significantly influencing the outcomes of porosity and pollutant tracking. The study revealed that a "Very High" suitability level enhances the likelihood of achieving the target outcome by a substantial multiplier of 34.0x. This underscores the critical role of suitability in decision-making processes, urging a strategic focus on factors contributing to a "Very High" suitability level.

Environmental factors, including geological and geophysical elements, positively impacted outcomes, while water surface, slope, and land use had negative influences. The actionable insights derived from this analysis recommend concentrating on positive geological and geophysical factors while addressing challenges associated with negative impacts through targeted interventions.

Sub-factors related to geology, geotechnical, geophysics, surface water, land use, and slope were found to play a crucial role in influencing outcomes. The study suggests enhancing porosity and pollutant tracking by prioritizing favourable sub-factors and addressing challenges associated with negative impacts.

Rating, percentage of influence, and sub-factor classification based on suitability in meters were identified as relatively important features. The study emphasizes the significance of considering higher ratings, higher

percentages of influence, and sub-factors falling under higher classes in decision-making processes to improve overall porosity and pollutant tracking.

The impact analysis of suitability levels, environmental factors, sub-factors, rating, percentage of influence, and class provided a nuanced understanding of their influence on porosity and pollutant tracking. Visual representations in figures 2 to 7 facilitated a comprehensive exploration of the relationships, aiding decision-makers in optimizing strategies based on environmental conditions.

The model fit analysis demonstrated a highly accurate and precise model, with a 100% accuracy rate and no incorrect predictions. The confusion matrix and corresponding metrics further validated the model's reliability, showcasing its effectiveness in predicting porosity and pollutant tracking outcomes.

In the prediction phase, the trained model successfully executed predictions on specific datasets, providing valuable insights and predictions tailored to each row. Four candidate sites were identified as highly suitable landfill sites, four as moderately suitable, and three as not suitable, based on the environmental criteria.

The deployment of the trained model in a real-world scenario showcased its applicability and effectiveness for similar investigations. The study's findings contribute significantly to the field of engineering geology, providing a robust framework for porosity and pollutant tracking, coupled with actionable insights for informed decision-making. Furthermore, four candidate sites were identified for moderately suitable landfill and other environmental purposes (Sokoto2, Maletе, Jimba and Omuaran) amongst several sites within the scope of this research while three sites are not suitable (sokoto1, Ijagbo and Oke oyi) due to their degree of porosity and other environmental factors.

REFERENCES

1. **Akinrinmade, A. O., Olasehinde, P. I., Olasehinde, D. A., Awojobi, M. O., Ige, O. O., & Olatunji, J. A. (202 C.E.).** Sanitary Landfill Sites Selection Using Multi-Criteria Decision Analysis And Gis-Modelling In Parts Of Kwara Sate, Nigeria. *IOSR Journal of Applied Geology and Geophysics (IOSR-JAGG)*, 8(5), 37–56. <https://www.iosrjournals.org/iosr-jagg/papers/Vol.%208%20Issue%205/Series-1/D0805013756.pdf>
2. **Alavi, N., Goudarzi, G., Jaafarzadeh, N., & Hosseinzadeh, M. (2012).** Municipal solid waste landfill site selection with geographic information systems and analytical hierarchy process: a case study in Mahshahr County, Iran. *Waste Management & Research*, 31(1), 98–105. <https://doi.org/10.1177/0734242x12456092>
3. **Al-Hanbali, A., Alsaaidеh, B., & Kondoh, A. (2011).** Using GIS-Based Weighted Linear Combination Analysis and Remote Sensing Techniques to Select Optimum Solid Waste Disposal Sites within Mafrag City, Jordan. *Journal of Geographic Information System*, 03(04), 267–278. <https://doi.org/10.4236/jgis.2011.34023>
4. **Allen, A. (2001).** Containment landfills: the myth of sustainability. *Engineering Geology*, 60(1–4), 3–19. [https://doi.org/10.1016/s0013-7952\(00\)00084-3](https://doi.org/10.1016/s0013-7952(00)00084-3)
5. **Attanasi, E. D., Freeman, P. A., & Coburn, T. C. (2020).** Comparison of machine learning approaches used to identify the drivers of Bakken oil well productivity. *Statistical Analysis and Data Mining*, 14(6), 536–555. <https://doi.org/10.1002/sam.11487>
6. **Bagchi, a. (1994)** *Design, Construction and Monitoring of Landfills. 2nd Edition, John Wiley & Sons, Inc., New York.* - References - Scientific Research Publishing. (1994). <https://www.scirp.org/reference/referencespapers?referenceid=1109785>. Retrieved December 19, 2023, from <https://www.scirp.org/reference/referencespapers?referenceid=1109785>
7. **Barzegar, R., Moghaddam, A. A., Deo, R. C., Fijani, E., & Tziritis, E. (2018).** Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms. *Science of the Total Environment*, 621, 697–712. <https://doi.org/10.1016/j.scitotenv.2017.11.185>

8. **Bell, F. G. (2000).** Geological Hazards: Their assessment, avoidance and mitigation. *Disaster Prevention and Management*, 9(3). <https://doi.org/10.1108/dpm.2000.07309cad.002>
9. **Bressan, T. S., De Souza, M. K., Girelli, T. J., & Chemale, F. (2020).** Evaluation of machine learning methods for lithology classification using geophysical data. *Computers & Geosciences*, 139, 104475. <https://doi.org/10.1016/j.cageo.2020.104475>
10. **BSI (1990) BS 1377 1990—Methods of Test for soils for civil Engineering purposes.** British Standards Institute, Milton Keynes. - References - Scientific Research Publishing. (n.d.). <https://www.scirp.org/reference/ReferencesPapers?ReferenceID=2044495>
11. **Chelani, A. B. (2009).** Prediction of daily maximum ground ozone concentration using support vector machine. *Environmental Monitoring and Assessment*, 162(1–4), 169–176. <https://doi.org/10.1007/s10661-009-0785-0>
12. **Engle, M. A., & Brunner, B. (2019).** Considerations in the application of machine learning to aqueous geochemistry: Origin of produced waters in the northern U.S. Gulf Coast Basin. *Applied Computing and Geosciences*, 3–4, 100012. <https://doi.org/10.1016/j.acags.2019.100012>
13. **EPA (2006). EPA Landfill Manuals Manual On Site selection Draft for consultation.** Environmental Protection Agency. - References - Scientific Research Publishing. (n.d.). <https://www.scirp.org/reference/referencespapers?referenceid=2606566>
14. **Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., & Strachan, R. (2014).** Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems With Applications*, 41(4), 1937–1946. <https://doi.org/10.1016/j.eswa.2013.08.089>
15. **Ferentinou, M., & Sakellariou, M. (2007).** Computational intelligence tools for the prediction of slope performance. *Computers and Geotechnics*, 34(5), 362–384. <https://doi.org/10.1016/j.compgeo.2007.06.004>
16. **Frempong, E. M. (1999).** Engineering geological assessment of a proposed waste disposal site in coastal southwestern Ghana. *Environmental Geology*, 37(3), 255–260. <https://doi.org/10.1007/s002540050383>
17. **Gurney, K. R. (2018).** An introduction to neural networks. In *CRC Press eBooks*. <https://doi.org/10.1201/9781315273570>
18. **Howard, A. D., & Remson, I. (1978).** *Geology in environmental planning*. <https://catalog.lib.kyushu-u.ac.jp/ja/recordID/1000976193>
19. **Hughes, K. L., Christy, A. D., & Heimlich, J. E. (2005a).** Landfill Types and Liner Systems. *OhioState Fact Sheet*. <https://www.ohioline.ag.ohiostate.edu>
20. **Hughes, K. L., Christy, A. N., & Heimlich, J. E. (2005b).** *Landfill types and Liners system*. <http://ce561.ce.metu.edu.tr/files/2013/11/liner-1.pdf>. Retrieved December 19, 2023, from <http://ce561.ce.metu.edu.tr/files/2013/11/liner-1.pdf>
21. **LandsAt Missions | U.S. Geological Survey. (2023, December 7).** <https://www.usgs.gov/landsat-missions>
22. **Leao, S., Bishop, I. D., & Evans, D. W. (2004).** Spatial–temporal model for demand and allocation of waste landfills in growing urban regions. *Computers, Environment and Urban Systems*, 28(4), 353–385. [https://doi.org/10.1016/s0198-9715\(03\)00043-7](https://doi.org/10.1016/s0198-9715(03)00043-7)
23. **Li, T. L. H., Chan, A. B., & Chun, H. W. (2010).** Automatic musical pattern feature extraction using convolutional neural network. *International MultiConference of Engineers and Computer Scientists*, 546–550. https://www.researchgate.net/profile/Antoni_Chan2/publication/44260643_Automatic_Musical_Pattern_Feature_Extraction_Using_Convolutional_Neural_Network/links/02e7e523dac6bb86b0000000.pdf
24. **Marjanović, M., Kovačević, M., Bajat, B., & Voženilek, V. (2011).** Landslide susceptibility assessment using SVM machine learning algorithm. *Engineering Geology*, 123(3), 225–234. <https://doi.org/10.1016/j.enggeo.2011.09.006>
25. **Michaelaschloegl. (n.d.).** *Weather ilorin.* Meteoblue. https://www.meteoblue.com/en/weather/week/ilorin_nigeria_6296447

26. **Mignan, A., & Broccardo, M. (2020).** Neural Network Applications in Earthquake Prediction (1994–2019): Meta-Analytic and Statistical Insights on their Limitations. *Seismological Research Letters*, 91(4), 2330–2342. <https://doi.org/10.1785/0220200021>
27. **Mohaghegh, S. D. (2020).** Subsurface analytics: Contribution of artificial intelligence and machine learning to reservoir engineering, reservoir modeling, and reservoir management. *Petroleum Exploration and Development*, 47(2), 225–228. [https://doi.org/10.1016/s1876-3804\(20\)60041-6](https://doi.org/10.1016/s1876-3804(20)60041-6)
28. **Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L., & Beroza, G. C. (2020).** Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-17591-w>
29. **Nathanson, J. A. (2007).** *Basic environmental technology: water supply, waste management, and pollution control*. <https://ci.nii.ac.jp/ncid/BB1747442X>
30. **NGSA. (2017).** <https://ngsa.gov.ng/>. <https://ngsa.gov.ng/geological-maps/>
31. **Osgof. (n.d.). HOME.** OFFICE OF THE SURVEYOR-GENERAL OF THE FEDERATION. <https://osgof.gov.ng/>
32. **Oweis, I. S., & Khera, R. P. (1990).** Geotechnology of waste management. *Choice Reviews Online*, 28(04), 28–2161. <https://doi.org/10.5860/choice.28-2161>
33. **Oyegun, R. O. (1985).** The use and waste of water in a third world city. *GeoJournal*, 10(2). <https://doi.org/10.1007/bf00150741>
34. **Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016).** A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, 2016(1). <https://doi.org/10.1186/s13634-016-0355-x>
35. **Saliu, O., Curilla, D., Lennon, M., & Chung, A. E. (2020).** Lessons Learned: Deep Learning for Mineral Exploration. *First EAGE Conference on Machine Learning in Americas, Volume 2020, pp.1–1, 1–1*. <https://doi.org/10.3997/2214-4609.202084021>
36. **Şener, B., Süzen, M., & Doyuran, V. (2005).** Landfill site selection by using geographic information systems. *Environmental Geology*, 49(3), 376–388. <https://doi.org/10.1007/s00254-005-0075-2>
37. **Smolic, H. (2023, September 12).** *No-code predictive Analytics for Data Teams | Graphite Note*. Graphite Note. <https://graphite-note.com/>
38. **Snodgrass, J. E., & Milkov, A. V. (2020).** Web-based machine learning tool that determines the origin of natural gases. *Computers & Geosciences*, 145, 104595. <https://doi.org/10.1016/j.cageo.2020.104595>
39. **US Environmental Protection Agency (EPA) (2005) National Management Measures to Control Non-Point Source Pollution for Urban Areas. Chapter 7 and 8, Document No. EPA 84 1-B-05-004, Washington DC. - References - Scientific Research Publishing. (n.d.).** <https://www.scirp.org/reference/referencespapers?referenceid=2270519>
40. **World Bank. (2017, November 22).** *Guidelines for selection and construction of sanitary landfills*. www.worldbank.edu.org. <https://www.worldbank.edu.org>
41. **WRSC-Westinghouse Savannah River Company. (1992).** *Preliminary site selection report for the new sanitary landfill at the Savannah River Site*. <https://doi.org/10.2172/10163617>
42. **Zhao, T., & Wang, Y. (2020).** Interpolation and stratification of multilayer soil property profile from sparse measurements using machine learning methods. *Engineering Geology*, 265, 105430. <https://doi.org/10.1016/j.enggeo.2019.105430>
43. **Zuquette, L. V., Palma, J. B., & Pejon, O. J. (2005).** Environmental assessment of an uncontrolled sanitary landfill, Pocos de Caldas, Brazil. *Bulletin of Engineering Geology and the Environment*, 64(3), 257–271. <https://doi.org/10.1007/s10064-004-0268-z>