Estimation of Soil Moisture and Earth Resistivity Using Wenner's Method and Machine Learning

by

Valdimiro Cussei

Bachelor of Engineering Science, The University of Western Ontario, 2019

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Applied Science

in the
School of Engineering Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Fall 2020

Declaration of Committee

Name:	**Valdimiro Cussei**
Degree:	**Master of Applied Science**
Title:	**Estimation of Soil Moisture and Earth Resistivity Using Wenner's Method and Machine Learning**
Committee:	**Chair:** **Shahram Payandeh** Professor, Engineering Science
	Bonnie Gray Supervisor Professor, Engineering Science
	Teresa Cheung Committee Member Assistant Professor of Professional Practice, Engineering Science
	Andrew Rawicz Examiner Professor, Engineering Science

Abstract

The present research consists of using Wenner's four electrodes method to measure the electrical resistivity of soil (e.g., clayey silt and clay), applying two machine-learning algorithms (k Nearest Neighbor (KNN) and Support Vector Machine (SVM)) to predict the type of soil. Such predictions may be leveraged, e.g., to extract parameters to help choose materials to withstand electrochemical corrosion in a hybrid environment (soil and moisture). A dataset of 162 sample points was obtained from the literature (151 training, 11 testing points). From laboratory experiments, 26 sample points (corresponding to 130 measurements) were obtained; 6 points were added to the literature training dataset, and 20 were used as testing points for final validation. The results show that given the electrical resistivity of soil and its moisture, the KNN model is capable of predicting the type of soil with accuracy, error rate, sensitivity, specificity, and precision of 70%, 30%, 64%, 83%, and 90% respectively. In contrast, the SVM presented an error rate and accuracy of 44.1% and 55.9 % respectively.

Keywords: soil resistivity; soil moisture, corrosion; k-NN (k-Nearest Neighbor); corrosion

I am truly grateful to my sisters for believing in me even when I doubted myself several times.

I dedicate the chapters of this enterprise to my parents and two brothers

Acknowledgments

I am thankful to my senior supervisor, Dr. Bonnie Gray, for guiding me throughout this project. Without your help, it would be extremely difficult to finish this thesis. I am also grateful to you for becoming my supervisor when I was left without one. Your dedication to helping me improve myself academically and professionally will never be forgotten.

I am grateful to my former English teacher, Olga Kharytonava, whose willingness to help me to succeed went beyond the classroom. Without your guidance, I would not be the man I am today. Because you challenged me intellectually, I was able to transform myself to become a better version of myself.

I would like to thank Dr. Teresa Cheung, Andrew Rawicz, and Shahram Payandeh for your guidance and support during these difficult times caused by the COVID-19 pandemic. I am grateful for finding time to support me and becoming part of my defense committee members.

It would be very difficult to finish this project without the help and emotional support of Yogiraj Kachhela, Joshua Ighalo, Alexandre Cachunjua, Kiran Sonea, Jasbir Patel, Edson W. Silva, and Bozena Kaminska.

Thank you all for your support.

Author

Valdimiro Cussei

Table of Contents

List of Tables

List of Figures

List of Acronyms

ACPS	Adaptive Corrosion Protection System
FN	False Negative
FP	False Positive
GUI	Guider User Interface
H_2SO_4	Sulfuric Acid
k-NN	k Nearest Neighbor
SFU	Simon Fraser University
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

Chapter 1.

1.1. Introduction

This research is part of a collaboration between Simon Fraser University and Powertech Labs, a division of BC Hydro, with the overall goal to provide instrumentation to help analyze corrosive environments and avoid corrosion of power grillage foundations. The larger project includes methods to reduce the cost of corrosion in power systems industries via a new miniature Adaptive Corrosion Protection System (ACPS) that protects a target metal by monitoring the corrosion status and minimize the protection parameters required. This dissertation presents a more preventive approach that aims to measure the electrical resistivity of soil using Wenner's four-electrode technique and use the measured electrical resistivity to predict the type of soil under investigation to help determine the potential corrosivity of a particular environment and aid engineers in recommending materials that may withstand corrosion in that specific environment.

The system proposed in this dissertation computes the resistivity of soil, predicts the type of soil, and leverages parameters such as soil electrical resistivity and moisture to aid engineers in the identification of potentially corrosive environments and selection of the best materials to withstand corrosion in that hybrid environment (soil and moisture). The advantage of focusing on prevention instead of protection is that it allows engineers to collect more information about the potential corrosion aggressiveness of the soil before placing the metal underground.

The environment considered in this dissertation is a hybrid environment of soil and moisture, which may cause grillage structures to undergo electrochemical corrosion. This type of corrosion is characterized by the destruction of the metallic structure caused by direct contact with the electrolyte solution (e.g. soil and moisture) that generates electrons that move from the anode to the cathode [1]. Electrochemical corrosion differs from chemical corrosion that is the redox in which the electrons of the metal are passed directly to the substances (water vapor or gas at high temperature) in the environment [1]. In this research, we intend to aid engineers in selecting the recommended material that

may withstand electrochemical corrosion when buried underground. This will be done by measuring the electrical resistivity of the soil and use it to select the recommended material that may withstand corrosion in that specific environment. The value of soil electrical resistivity indicates the relative ability of a medium to carry electrical current. It also influences the degree of corrosion in an underground metallic structure and the growth of agricultural products. Testing the resistivity of soil has become an important step of soil analysis before construction and plantation in Civil engineering and agriculture respectively [2]. However, measuring soil resistivity is not an easy task because several factors affect the electrical resistivity of soils, such as mineral composition, grain size, porosity, and organic materials [3]. Electrical resistivity is defined as the electrical resistance measured between two opposite faces of a unit cube [4]. However, the resistivity of soil depends on many things, including the layer being measured, soil type, moisture content, grain size, the closeness of packing, temperature, chemical composition, salt concentration, season, etc. [5].

Several techniques of measuring soil resistivity have been developed to better understand its influence in the degradation of metallic materials and to protect vegetation. Four techniques of measuring soil resistivity are summarised in this dissertation namely, Wenner's four electrodes, Schlumberger's technique, Boy's method, and Multispectral Imagining technique discussed in [4],[6], and [5], respectively. However, only Wenner's method is utilized during the laboratory experiment because it is effective, accurate, and easy to implement [7]. The resistivity and soil moisture measured from the laboratory setup is input to a k-nearest neighbor and support vector machine algorithm to predict the type of soil under investigation. The parameters measured and estimated through the machine learning algorithms are then leveraged to select the best material for a specific environment to avoid or delay corrosion.

Many scholars have investigated the relationship between resistivity, soil moisture, and temperature. Recent reports have stated that there is a non-linear interdependency between resistivity and soil moisture [8]. Vivek Sai et al. conducted a study wherein electrical resistivity was measured using Wenner's four-electrode techniques to observe the credibility of the existing computational methods for simulating soil moisture content

from the resistivity of the soil [8]. After calculating the electrical resistivity and extracting soil moisture from the resistivity measurements, they verified the uncertainty of the results by comparing the actual soil moisture (measured by a moisture sensor) to the estimated value (calculated using a regression equation) [8]. Since the discrepancy between the estimated soil moisture value and the measured value was very small, the study concluded that soil moisture can be extracted from the measured electrical resistivity [8].

It is important to note that the electrical resistivity of a given soil at a constant temperature, and water content might vary slightly depending on the depth, and separation of the electrodes being used. Liangfu et al. reported in [9] that the reason for soil conductivity variation at different depths is the change of the environment resistivity due to the increase of active ions within that area. However, the resistivity of the soil can be estimated by taking an average of the electrical resistivity from different depths, and locations within a short radius.

This technique reduces the time and cost when investigating geotechnical parameters of difficult terrain. For example, the usage of heavy equipment to measure different parameters in terrain with difficult access becomes unnecessary if a resistivity meter can be used instead. Abidin et al. conducted a laboratory study wherein a resistivity meter was used to calculate the resistivity of soil and the density of the soil was extracted using a regression equation [10]. The bulk density's regression coefficient (R^2) was estimated to be 0.7016. Although the value for R^2 is not close enough to 1, this value could be improved by using more accurate resistivity meters and soil moisture sensors.

The state of the art in the investigation of soil characteristics has been advancing in the last few years from only measuring the resistivity of soil to predict the soil type, moisture, and drainage using machine learning algorithms [8], [9], [10], [11], [12], [13], and [14]. In terms of machine learning, several scholars have been applying diverse machine learning algorithms to predict natural soil drainage, and soil properties. For example, in [12] a boosted tree algorithm was used to predict natural soil drainage properties of different regions. This study aimed to classify different regions based on

their natural soil drainage. Another research conducted in [11] used a Support Vector Machine (SVM) algorithm to estimate the type of soil under investigation based on soil features such as backscatter and incident angle from tropical rainfall. In addition to the two studies mentioned above, the research presented in [13] used an Artificial Neural Network (ANN) to create a digital mapping of different types of soil. The ANN algorithm leverages soil properties and locations to predict different types of soil. The method presented in this thesis differs from these previously described methods by combining a machine learning approach (k Nearest Neighbor approach) and Wenner's four electrodes technique and by using parameters such electrical resistivity and soil moisture to select materials less likely to corrode in a given environment.

One of the most used devices to measure soil resistivity is the AEMC 4630 Rechargeable Digital 4-Point Ground Resistance Tester. This device measures the resistivity of soil using Wenner's technique [14]. The other device that is being currently developed is the Adaptive Corrosion Projection System (ACPS). The ACPS is a current-sourced device that protects a target metal using a cathodic protection technique. Additionally, it monitors the state of corrosion while optimally protecting the target structure. However, this device is still under development in a joint collaboration between Simon Fraser University and Powertech.

The device proposed in this thesis is part of a larger project with BC Hydro and Powertech Labs ACPS project to develop instrumentation that helps to analyze and address corrosion for power grillage structures. The machine-learning algorithms presented in this thesis could supplement the ACPS by aiding engineers to predict different types of soil. This additional feature to the ACPS would result in a more robust and updated device to the new artificial intelligence technology trending today. If implemented, the proposed device could help fill a gap in the current state of the art as part of the equipment that measures electrical resistivity, soil moisture, and predicts the type of soil under investigation.

1.2. Contribution of the thesis work

The utilization of Wenner's method to measure electrical resistivity is the current state of the art in the field of geology. However, neither currently available commercial systems nor academic research has explored the combination of Wenner's technique together with machine learning algorithms for soil type prediction, and leveraged electrical resistivity to predict the type of soil under investigation to help determine the potential corrosivity of a hybrid environment (e.g. soil and moisture). The thesis research combines Wenner's technique and machine learning algorithms to measure electrical resistivity, soil moisture, and predict the type of soil under investigation. The results could then be potentially used to help engineers in recommending materials that may withstand corrosion in a hybrid environment (soil and moisture). Current systems do not have the capability of predicting the type of soil using a machine learning algorithm. One example of such a device that is commercially available is the AEMC 4630 Rechargeable Digital 4-Point Ground Resistance Tester. Although it measures the electrical resistivity of soil just like the proposed device, it cannot predict the type of soil, nor extract important parameters (such as electrical resistivity, and soil moisture) using a machine learning algorithm (k Nearest Neighbor (kNN)), which could be used to select material that may withstand electrochemical corrosion of underground metallic structures. In addition, such a combination of techniques aimed towards materials selection does not exist in the literature. The integration of machine learning into a Wenner's method tester would provide such resistivity testers the ability to predict the type of soil and store important information about the soil under investigation that could then be used to understand the characteristics of different types of soils in the future. In other words, the device could learn important soil characteristics whenever utilized in the field, which could then add to its prediction accuracy. For example, such an instrument may take the form of a portable hardware device and work with the ACPS to perform soil investigation such as measuring the electrical resistivity of soil and predicting different types of soil. These characteristics learned during field tests are dynamically utilized to improve the performance of the machine learning algorithms, and soil prediction over time

This addition of machine learning to such devices also provides the capability of engineers to utilize the same device to extract the resistivity of soil, soil moisture, and type of soil to select the best material that may withstand electrochemical corrosion within a hybrid environment (soil and moisture). The selection of the optimal material that may withstand corrosion in a hybrid environment would be determined in two steps: 1) the electrical resistivity and moisture would be extracted from the device measurement; 2) the extracted parameters would then be used to estimate the corrosion aggressiveness of the soil using corrosion standards such as the American Water Works Association (AWWA) for Ductile-Iron Pipe Systems [15]. Although machine learning (Support Vector Machine (SVM)) has been used for soil prediction in [12] by using features such as backscatter and incident angle from tropical rainfall to predict different types of soils, it has not to the author's knowledge, been used in combination of Wenner's four electrodes technique to extract parameters such as soil electrical resistivity, and moisture to delay corrosion by selecting the ideal material that may withstand corrosion within a hybrid environment (soil and moisture). The benefit of using such a device is that it allows engineers to be able to measure electrical resistivity (using Wenner's four electrodes technique), predict the type of soil under investigation (using a machine learning algorithm), and extract parameters to help in the selection of material that will better withstand electrochemical corrosion from a single electronic device. There is no device available in the market today or described in the scientific literature that is capable of performing all these three tasks from a single measurement.

1.3. Thesis outline

This dissertation is divided into several chapters. Each chapter of this thesis provides background information on the next section to help the reader fully understand the content in the subsequent sections. The chapters of this thesis are organized as follows:

- **Background (Chapter 2):** This chapter presents several reviews of different scholars' publications on the topic of corrosion and machine learning used in soil prediction research. A detailed explanation of the concept of corrosion, the different types of corrosion, and their impact on

the surface of the metal are presented. Also, an overview of the different types of soil and their properties are presented in this chapter. The influence of temperature and soil moisture on different types of soil is also discussed here. The chapter concludes with a discussion of the current state of the art in machine learning algorithms used to predict different types of soil.

- **Proposed Methodology (Chapter 3):** The proposed approach of this dissertation and the methodology utilized to implement it is outlined. The diagram of the proposed approaches is discussed. Also, the k-NN and SVM algorithms are defined in detail. Finally, the results of the implemented models are presented and discussed.

- **Preliminary experimental verification (Chapter 4):** In this chapter, the laboratory experiment setup, the devices used for the laboratory experiment, and the techniques utilized are presented. The results obtained from the models and laboratory experiments are reported in detail.

- **Limitations (Chapter 5):** Several limitations that might have affected the result obtained during the laboratory experiment are reported in this section. The implication of these limitations to the overall conclusion of this dissertation are outlined.

- **Conclusions and Future Work (Chapter 6):** The outcomes from the thesis research are presented and conclusions are reported in the context of combined corrosion and machine learning research. The results and implications of the k-NN and SVM algorithms are outlined. Future work to improve the results obtained from this research is also discussed in detail.

Chapter 2.

Background

This chapter presents a background on the corrosion process, soil types, and measurement techniques. The chapter concludes with a discussion of machine learning algorithms currently being used to perform the prediction of different types of soil and properties.

2.1. Basic concepts of corrosion

There are several definitions of corrosion depending on the purpose of the research or the field of study. For this research, we define corrosion as "an electrochemical reaction between a material, usually a metal, and its environment that produces a deterioration of the material and its properties" [16]. The electrochemical reaction is one in which a metal loses electrons by transferring them to the environment and undergoes a valence change, thus becoming positively charged with a value **z.** The environment of a metal corroding is everything that surrounds the material. According to Joseph et al [16], these environments are classified as a physical state (gas, liquid, or solid), chemical composition (constituents and concertation), and temperature. Besides, some environments are hybrid soil-liquid (this is our environment of interest). Since these environments have their conductivity, they exchange electrons, we call them electrolytes. Conductive solutions such as an electrolyte contain positively and negatively charged ions called cations and anions, respectively. An example of a corrosion reaction process is illustrated below. In this reaction extracted from [17], a metallic material is immersed within a sulfuric acid ($\mathbf{H_2SO_4}$) solution wherein metal oxidation (the loss of one or more electrons by the metal) occurs through an anodic reaction and reduction (the metal gaining one or more electrons) through a cathodic reaction.

$$\mathbf{M \rightarrow M^{+z} + ze^{-}\ (Anodic \equiv Oxidation)\ (eq.\ 2.1a)}$$

$$\mathbf{zH^{+} + zSO^{-}_{4} + ze^{-} \rightarrow \frac{z}{2} H_2SO_4\ (Cathodic \equiv Reduction)\ (eq.\ 2.1b)}$$

$$\mathbf{M + zH + zSO^{-}_{4} \rightarrow M^{+z} + \frac{z}{2} H_2SO_4 \ (Overall \equiv Redux) \ (eq.\ 2.1c)}$$

In the chemical reactions described in eq.2.1a, eq.2.1b, and eq.2.1c M, H^+, and z are Metal, Hydrogen cation, and Valence or Oxidation State respectively.

The rate at which metal corrodes depends on several factors such as temperature, type of metal involved, soil moisture (in a hybrid environment), soil composition, diffusion, conductivity, type of ions, pH value, electrochemical potential, and more. Therefore, to properly understand corrosion, it is imperative to identify the environment in which the material is to be exposed. Also, the material that corrodes needs to be identified to understand the corrosiveness or aggressiveness of an environment on that material [16].

Corrosion in a metallic structure may be manifested in different forms. In terms of classification, there are several types of corrosion namely: general corrosion, localized corrosion, atmospheric corrosion, and galvanic corrosion.

2.1.1. General corrosion

In this form of corrosion, the compromised surface area of the metal/alloy is usually corroded completely. The exposed surface area is submersed within the environment which could be a liquid electrolyte (chemical solution, a liquid metal), gaseous electrolyte (air, CO_2, $SO^{-}_{2,}$ etc.), or a hybrid electrolyte (solid and water, biological organisms, etc.) [17]. The nature of the corrodent which causes general corrosion could be either wet (the electrolyte could be a liquid or moisture), dry (it usually involves reaction with high-temperature gases), or both.

2.1.2. Localized corrosion

This form of corrosion usually occurs on a specific surface area of the exposed metal. Localized corrosion is very difficult to control compared to other forms of corrosion [17]. In terms of classification, the localized corrosion can be classified as crevice corrosion (associated with a stagnant electrolyte such as dirt corrosion product, sand, and more), filiform corrosion (a type of crevice corrosion that occurs under a protective film), pitting

corrosion (corrosion that causes destructive pits), oral corrosion (occurs on dental alloys exposed to saliva), and biological corrosion (caused by fouling organisms) [17].

Additionally, localized corrosion might possess a macroscopic and microscopic form. Microscopic localized corrosion attack could cause considerable damage (or lead to structure failure) before the corrosion becomes available to the naked eyes. On the other hand, macroscopic localized corrosion attack is visible with the naked eyes or becomes visible when viewed with a low power magnifying device [16].

2.1.3. Atmospheric corrosion

In this form of corrosion, the entire metal surface area exposed to the corrosive environment is converted into its oxide form (only if the metallic material has a uniform microstructure) [17]. Atmospheric corrosion is usually a uniform and general attack phenomenon that manifest in different forms. For example, in uniform atmospheric corrosion, a brown-color corrosion layer (ferric hydroxide compound known as rust) could develop on a compromised steel surface. In contrast to localized corrosion, this form of corrosion is usually visible to the naked eyes.

2.1.4. Galvanic corrosion

This type of corrosion attack occurs when two dissimilar electrodes/metals are connected through an electrolyte environment leading to either a chemical or electrochemical reaction in which current flows from more negative metal to the more positive potential metal. The transfer of electrons from the more negative potential (anode) to the more positive (cathode) metal causes the anodic surface area to oxidize, thus leading to corrosion. This form of corrosion is easily prevented by proper corrosion design. For example, reference [17] states that “in selecting two metals or two alloys for a galvanic coupling, both metals should have similar potential or should be placed close to each other in series to suppress galvanic corrosion”. Therefore, the higher the difference of standard potentials of two coupled metals, the more enhanced galvanic corrosion becomes.

2.2. The economic impact of corrosion

Recent studies reported that the annual direct cost of corrosion to the industrial economy is approximately 3-4% of the country's Gross National Product (GNP) [18], [19]. For example, the United States of America spends approximately $276 billion every year on corrosion-related damage [18]. It is thus paramount to reduce the amount of money spent on corrosion. Industries are not the only institutions that lose exorbitant amounts of money due to corrosion; it affects everyone's daily life as well. For example, manufacturers raise the price of customers' products due to the high cost of machine maintenance. Also, corrosion is one of the causes of products (e.g. oil) spills and pollution that affect people's wellness.

There are two types of costs related to corrosion namely direct and indirect costs. Reference [19] defines direct cost as losses that can be quantitatively accounted for such as replacement cost, protection cost, corrosion inhibition, research, and development. On the other hand, indirect costs are losses that cannot be quantitatively evaluated such as loss of product to spill and fire, loss of revenue due to downtime, loss of efficiency of equipment, contamination of products, environment pollution, etc.

2.3. Soil characteristics

To understand soil conductivity, it is pivotal to know some of the characteristics and constituents of the terrain being investigated. These constituents influence the conductivity of soil because of conditions such as temperature and saturation level (water content). Soil consists basically of the following components: mineral material (clay, silt, and sand), organic material, water, and gases. In terms of mineral material, soil can be classified according to sand, silt, and clay size range. These three components can be identified according to their diameters. The diameters of sand, silt, and clay are 0.05-2 millimeters, 0.002-0.005 millimeters, and less than 0.002 millimeters, respectively [3]. Reference [3] also states that the content of clay in soil affects the soil conductivity, thus different soil types might have different conductivities.

Different types of soil may combine to create a new form of soil. The chemical constituents involved in the formation of soils affect their resistivity and their range of conductivity. The conductivity of soil is determined by porosity, moisture content, the concentration of dissolved electrolytes in the contained moisture, temperature and phase state of the pore water, and amount and composition of colloids [3]. This is because conductivity is electrolytic and takes place through the moisture-filled pores and passages contained within the insulating matrix [3]. On the other hand, the electrical resistivity of soil changes with respect to moisture (water content), temperature, and more [20].

2.4. Measurement techniques

Several techniques of measuring soil resistivity have been developed to better understand its influence in the degradation of metallic materials and to protect vegetation. We present a review of four techniques of measuring soil resistivity namely the Wenner's method, Schlumberger's technique, Boy's method, and Multispectral Imagining technique.

2.4.1. Wenner's four-electrode method

This is one of the best techniques used to measure soil resistivity because it is simple to implement compared to other testing methods [7]. When the soil is contained within a box, in Wenner's method, the relationship between the probe separation and the depth of penetration depends on the standard being used. The World Trade Organization Technical Barriers standard requires that four electrodes are placed with equal separation in a straight line in the surface of the soil to a depth of not more than 5% of the minimum separation of the electrodes [4]. The separation of the electrodes is chosen according to the soil strata of interest (Figure 2.1) so that the measured resistivity represents the average resistivity of a hemisphere of the soil of a radius equal to the electrode separation [4]. However, the AEMC Instrument [14] and [6], requires that the depth of penetration of the electrode be less or equal to 50% of the electrode separation (less or equal to half of the electrode separation).

For the laboratory measurements, the depth of penetration of the electrodes (e.g., 1, 2, or 4 cm) was approximately half of the separation (2, 4, or 8 cm, respectively) of the electrodes. A voltage is applied to the outer electrodes causing a current flow in the electrodes. The injected current (which flows radially outwards from its point source) generates a current density in the ground which is related to the electric field that creates the voltage drop [21]. In Wenner's four electrodes method, the two inner probes measure the electrical potential (voltage drop) caused by the variation in electrical conductivity underground (which results from the current flow) [21]. Box 1 was clay silt and was used for electrode penetration depths of 1 cm and 2 cm. Box 2 was clay and used for electrode penetration depths of 1, 2, and 4 cm. For the laboratory experiment, the volume conductor of soil in boxes 1 and 2 were 3689 cm^3 (clay silt) and 11722 cm^3 (clay). The depth of soil in box 1 was 8cmwhile the depth of soil in box 2 was 11cm. Based on the guideline for Wenner's method measurements, the box depths are considered sufficient [4]: their depths are each at least 3-4 times the depth of penetration of the electrodes.

The measured voltage and the applied current are used to calculate the mean resistance of the soil sample [22] (**eq. 2.3a**). If the current-carrying electrodes are not spaced equally as the potential-measuring electrodes, the resistivity (ρ) is given by **eq.2.3b**. Otherwise, the soil resistivity is given by **(eq2.3c)** [4]. If the experiment is conducted in the laboratory wherein the soil is contained within a soil box**,** the resistivity is calculated by **eq. 2.3d.**

$$R = \frac{V}{I} \qquad \textbf{(eq.2.3a)}$$

$$\rho, \Omega \cdot \text{cm} = 95.76 * \text{b} * \text{R}/(1 - \frac{\text{b}}{\text{b+a}}) \qquad \textbf{(eq2.3b)}$$

where: b = outer electrode spacing, ft,

a = inner electrode spacing, ft, and

R = resistance, **Ω**.

$$\rho, \Omega \cdot \text{cm} = 2\pi\text{a} * \text{R (a in cm)} \qquad \textbf{(eq. 2.3c)}$$

where: a = inner electrode separation, and

R = resistance, **Ω**.

$$\rho, \Omega \cdot \text{cm} = \text{R} * \frac{\text{A}}{a} (\text{a in cm}) \qquad \textbf{(eq. 2.3d)}$$

where:

R = resistance, **Ω**,

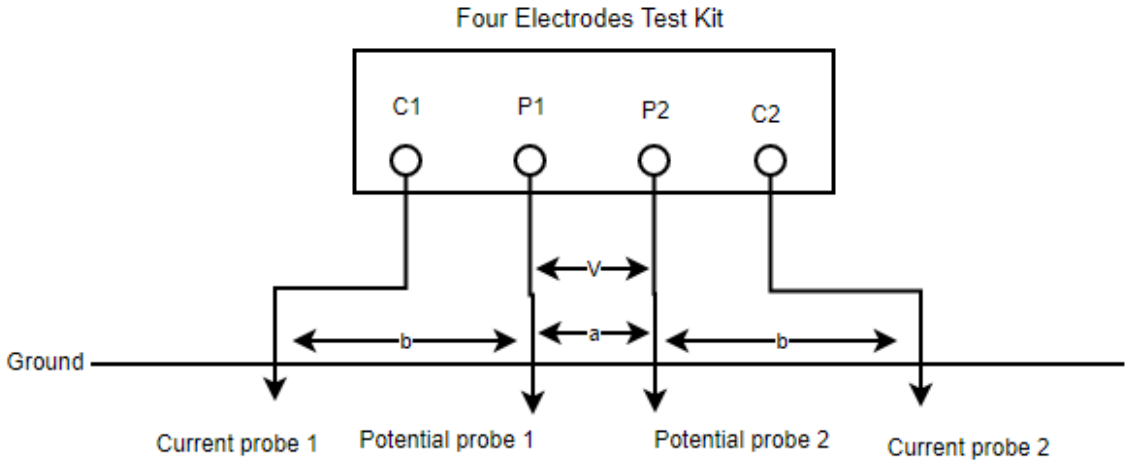

Figure 2.1. Wenner's Arrangement

Equations 2.3b and 2.3c were developed by the World Trade Organization Technical Barriers as a standard when measuring electrical resistivity of a given soil contained within a box [4]. In terms of boundary conditions, the outer electrode spacing (b) and inner electrode separation (a) should always be greater than zero. For equation 2.3b, if the outer electrode spacing is equal to zero, the electrical resistivity will be zero also. On the other hand, if inner electrode separation is zero the multimeter shows a continuous increase in the electrical resistivity value (infinite electrical resistivity as observed during the laboratory experiment). Additionally, as the inner electrode separation (b) increases the electrical resistivity decreases compared to a smaller value of a.

From equations 2.3b and 2.3c, the electrical resistivity depends on a and b, which are the inner electrode separation and outer electrode spacing, respectively [23]. Also, equations 2.3c can be used to calculate the electrical resistivity of a given soil regardless of the

actual placement of the electrodes on the surface of the soil [23]. The electrical resistivity of the laboratory soil was measured as instructed by the World Trade Organization Technical Barriers standard. This technique is widely used to measure the electrical resistivity of soil contained within a box [4].

For the laboratory experiment, we used four electrodes to measure the electrical resistivity of the soil instead of two electrodes. The usage of 4 electrodes instead of two electrodes to measure the electrical resistivity of a given soil is preferred to avoid unpredictability and measurement errors related to using only two electrodes [24]. When using two electrodes to measure the electrical resistivity of a given soil, the contact resistivity between the electrode and the soil is also added to the soil electrical resistivity. Thus, it is not a good technique to use wherein we only intend to measure the electrical resistivity of the soil sample. Wenner's four-electrode technique has been long preferred over two electrodes to avoid the unpredictability and measurement errors related to the latter method since 1931 [24].

2.4.2. Schlumberger's soil resistivity testing method

Schlumberger's soil resistivity testing has the same arrangement as Wenner's method in Figure 2.1. However, the resistivity of soil in this technique is measured differently wherein the inner and outer electrodes are not spaced equally. The inner electrodes (voltage probes) have the same distance with respect to the center, but different dimensions with respect to the outer electrodes Figure 2.2**.** According to B. Philip in [7], there are alternative techniques used to measure soil resistivity within Schlumberger's method, wherein the most used leaves the voltage probes stationary, while shifting the current electrodes out. This technique is different than Wenner's method where the test center is maintained so that all four electrodes have the same center [7].

In terms of human resources required to perform the measurement, Schlumberger's method is more economical than Wenner's technique since only the outer electrodes are moved. Additionally, the two current electrodes (outer electrodes) can be moved four or five times for each move of the inner electrodes (voltage electrodes) [6]. The soil resistivity is measured using **eq. 2.3e** [6]:

$$\rho, \Omega \cdot \text{cm} = \frac{\pi L^2 R}{2l} \qquad \textbf{(eq. 2.3e)}$$

where: L= distance from the center to the outer probe

l= distance to the center from the inner probe

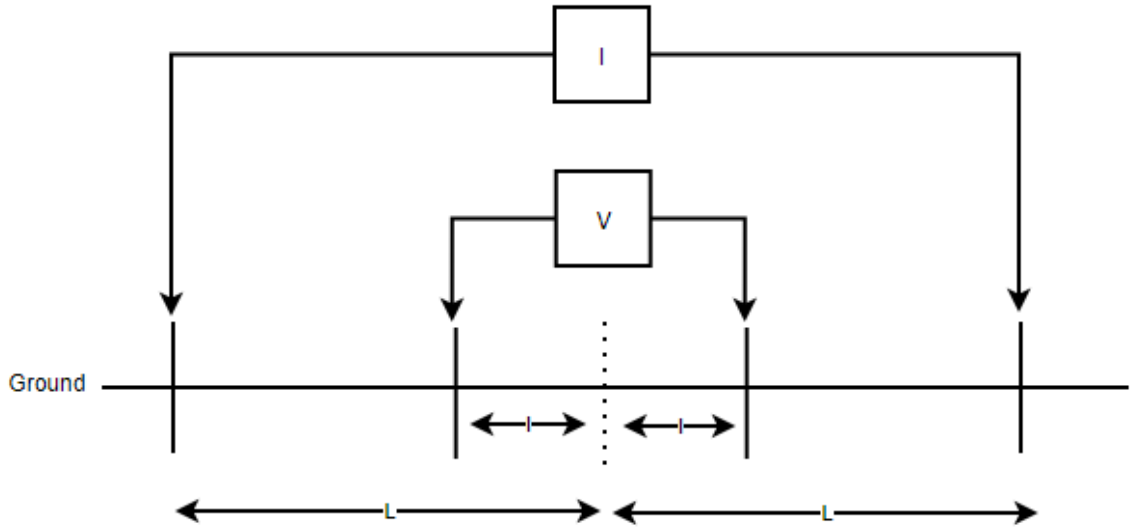

Figure 2.2. Schlumberger Arrangement

2.4.3. Boy's method for resistivity measurement

In this technique, the two current electrodes and a voltage electrode are fixed. The fourth electrode is shifted between measurements (starting close to the current probe) to determine the soil resistivity Figure 2.3**.** The Boy's method measures the resistivity around the outer stationary current probe at the movable voltage probe end [7]. In terms of time efficiency, the Boy's method is more efficient compared to Wenner's and Schlumberger's techniques. The Soil resistivity is measured using **eq. 2.3f:**

$$\rho = \frac{2(\text{Rmeasured} - \text{Rreference})\, \pi Z(W-Z)}{W-2Z} \qquad \textbf{(eq. 2.3f)}$$

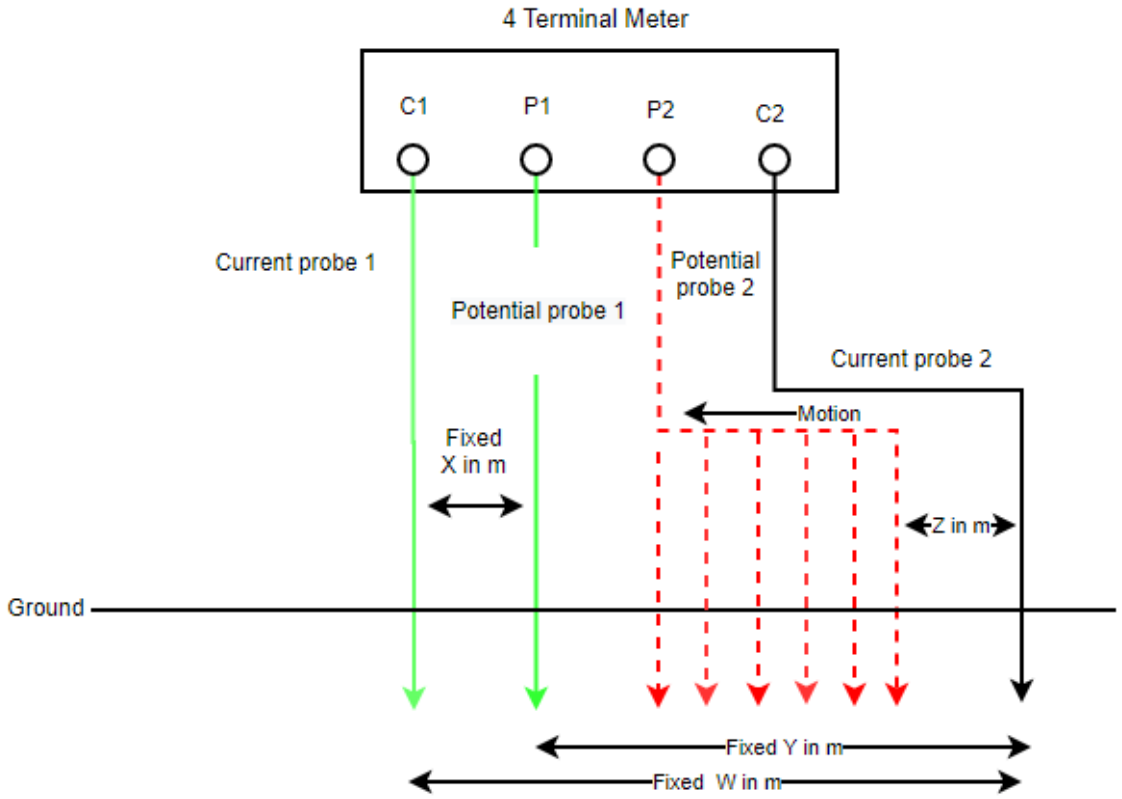

Figure 2.3. Boy's Method Arrangement

2.4.4. Multispectral Imagining method

This technique uses a statistics algorithm (maximum likelihood) and satellite (LANDSAT-7) images to approximate the soil electrical resistivity value. The maximum likelihood algorithm determines classes that maximize the probability of the likelihood of a sample [5]. By using the LANDSAT-7 image database of different types of soil and its resistivity, the system can predict the resistivity of an unknown soil. Although this is a powerful method of measuring soil resistivity, the results obtained are less accurate compared to the techniques mentioned previously. The inefficiency of this technique is caused by approximation errors introduced by the imaging system and the likelihood algorithm itself.

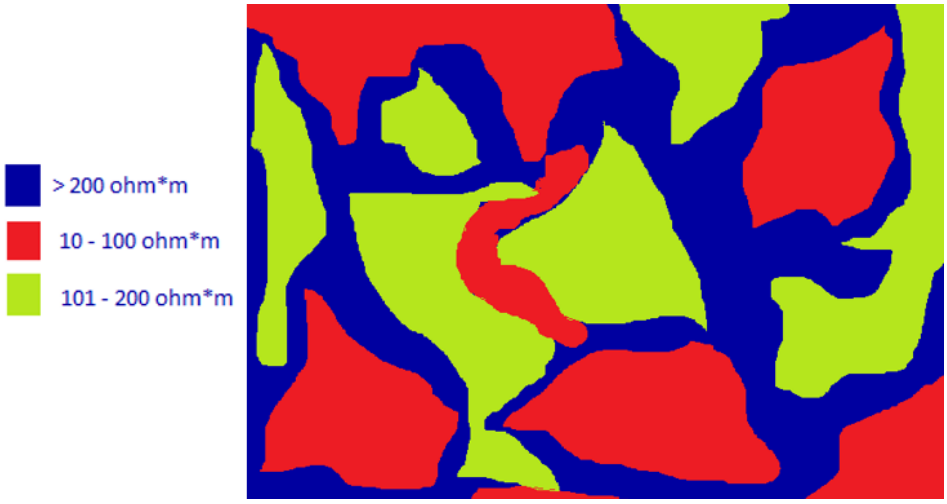

Figure 2.4. Earth resistivity map [adapted from 5]

The color map showed in Figure 2.4 was created using ENVI software wherein red, green, and dark blue indicates the soil resistivity [5]. In Figure 2.4**,** the areas in red have the lowest resistivity while those marked by dark blue have the highest resistivity. On the other hand, Figure 2.5 shows the flowchart of soil resistivity estimation wherein a supervised learning algorithm is used to predict the resistivity of different types of soil. The training and testing data of this algorithm were obtained using Wenner's method.

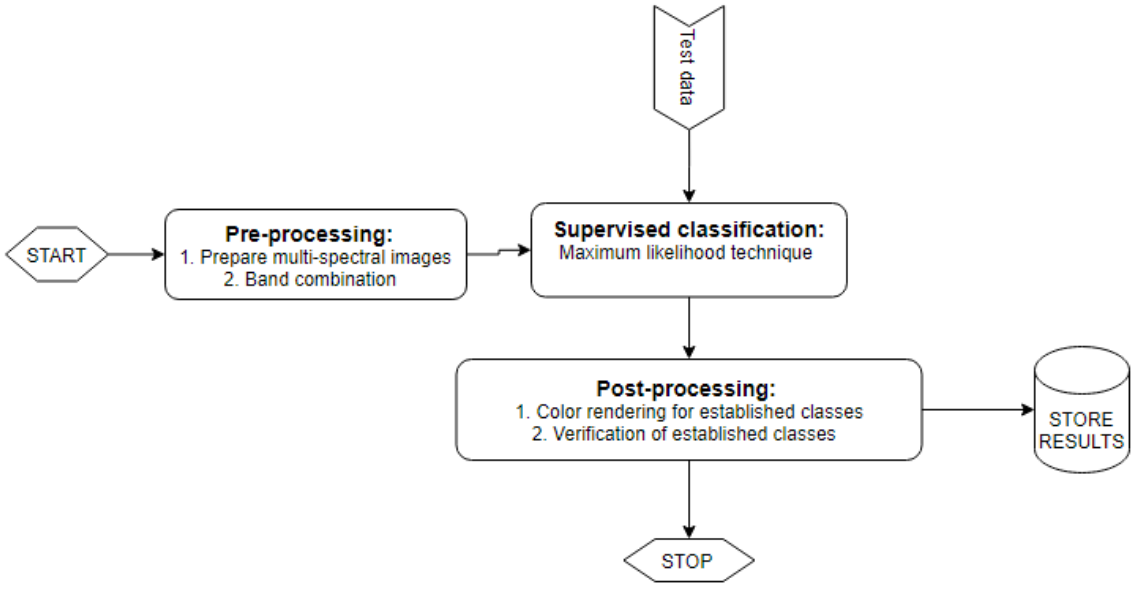

Figure 2.5. Flowchart for earth resistivity estimation using supervised learning [adapted from 5]

2.5. Prior Art in Machine Learning Applied to the Prediction of Types of Soil and Properties

Lemercier et al. conducted a study in Brittany (Northwestern France) to predict soil properties such as soil parent material and drainage. The soil parent material included bedrock formation and superficial deposit. A boosted classification tree and a two-step approach were employed as the algorithms used to predict the soil properties. The two steps approach consisted of predicting the parent material (PM) and leverage the PM as a predictive variable to estimate natural soil drainage [12]. In addition to the PM, environmental data representing known soil-forming factors such as terrain attributes (elevation, slope, profile and plant curvatures, sub-watershed hillslope length, hydrological distance from the nearest stream, aspect, relative elevation above the nearest stream, geological data, etc.) and landscape data derived from remote sensing data were used as predictive features [12]. The study concluded that "based on 20, 000 randomly pixels from the training area, selected PM and soil drainage were predicted with overall accuracies of 73 and 70% respectively". Also, the PM was calculated to be the most relevant variable to predict soil drainage.

Ahmad et al. researched the estimation of soil moisture using remote sensing data. A Support Vector Machine (SVM) learning technique was used to predict soil water content. The experiment was conducted in 10 sites in the Lower Colorado River Basin located in the western United States. The features used to predict the data are backscatter and incidence angle from Tropical Rainfall Measuring Mission (TRMM), and Normalized Difference Vegetation Index (NDVI) from Advanced Very High-Resolution Radiometer (AVHRR) [11]. The model was trained and tested with 5 years (from 1998-2002) and 3 years of data (from 2003-2005) [11]. After comparing the SVM model to Artificial Neural Network and Multivariate Linear Regression (MLR) model, the study concluded that the SVM model performs better for soil moisture estimation than ANN and MLR models.

A study conducted by Bodaghabadi et al. leveraged Artificial Neural Network to perform digital soil mapping. The area of interest of this study enclosed approximately 1 000

hectares and was located in the Borujen region of Chaharmahal-Va-Bakhtiari Province, Central Iran [13]. The study consisted of collecting several soil profiles of different types of soils and extracting its properties to feed into an ANN algorithm. The ANN algorithm was trained afterward using the soil properties extracted from the collected samples. Bodaghabadi et al. concluded that the overall accuracy of the ANN algorithm used was 50% and that the algorithm was capable of making an accurate prediction of the D (Fine, mixed, active, mesic Typic Calcixerepts), and F(Loamy-skeletal, carbonatic, mesic Typic Calcixerepts) soils series. However, the prediction reported for the A (Clayey-skeletal, carbonatic, mesic Petrocalcic Calcixerepts), C (Fine, carbonatic, mesic Typic Calcixerepts) and B (Fine, carbonatic, mesic Petrocalcic Calcixerepts), E (Fine-loamy, carbonatic, mesic Typic Calcixerepts), and G (Fine, mixed, active, mesic Calcic Haploxeralfs) soil series were acceptable and unacceptable respectively [13].

The AEMC 4630 Rechargeable Digital 4-Point Ground Resistance Tester is an electronic device that measures electrical resistivity using both Wenner's and Schlumberger techniques. Many engineers today use this electronic device to make sub-surface geophysical surveys for diverse soil investigation, protect metals against corrosion, and design grounding systems [14]. In terms of corrosion protection, the inverse relationship between soil resistivity and corrosion activity is used to prevent corrosion of underground pipelines [14]. The relationship between soil resistivity and corrosion is that "a decrease in resistivity relates to an increase in corrosion activity" [14]. Base on the fact that the most economic grounding installation is achieved at the location where the soil resistivity is the lowest, the AEMC 4630 Rechargeable Digital 4-Point Ground Resistance Tester is used for designing grounding systems [14]. This device has some similarities to the device proposed in this dissertation because they both measure electrical resistivity for sub-surface geophysical surveys and can be used for grounding installation.

Except for the AEMC 4630 Rechargeable Digital 4-Point Ground Resistance Tester, the studies presented in this section leverage machine learning algorithms to predict soil properties and mapping different types of soils. Although they all use machine learning algorithms such as SVM, ANN, and boosted tree, these studies were conducted for different purposes rather than implementing an electronic device. The three studies

presented require several data collections and soil profile analyses such as investigating soil elevation, soil moisture, slope, plant curvatures hillslope length, etc. The soil investigation conducted in each one of the studies presented here is time-consuming and expensive because they require more than a technician and several measurement tools to properly analyze the soil under investigation.

Although the AEMC 4630 Rechargeable Digital 4-Point Ground Resistance Tester and the proposed device measure electrical resistivity, perform sub-surface geophysical surveys and grounding systems design, they have several differences. The AEMC 4630 Rechargeable Digital 4-Point Ground Resistance Tester lacks the ability to perform soil prediction using a machine learning algorithm. Thus, it can merely be used to conduct soil electrical resistivity measurement. However, the device proposed in this dissertation will help engineers not just to perform soil electrical resistivity measurements, but predict the type of soil under investigation using a machine-learning algorithm, select the recommended material that may withstand corrosion, and measure soil moisture. In short, the proposed device can be used to extract parameters (such as electrical resistivity and soil moisture) to select material that may withstand electrochemical corrosion, dynamically leverages a machine-learning algorithm to improve its performance, and predict the type of soil under investigation. The implementation of a resistance tester that can help engineers to extract parameters (such as electrical resistivity and soil moisture) to select material that may withstand electrochemical corrosion, and predict the type of soil under investigation using machine learning would advance the current state of the art by providing an integrated combination of cutting edges techniques that have not been previously integrated.

Chapter 3.

Proposed Methodology

The measurement techniques reviewed in Chapter 2 are capable of computing soil resistivity with a certain accuracy. However, more improvements are needed to increase the measurement time efficiency and accuracy when testing non-homogeneous soil and to decrease the errors caused by the continuous motion of probes. For instance, testing soil resistivity using Wenner's and Schlumberger's methods is time-consuming and the user operating the device needs to move the probes around several times. Unless the measurement is being performed by several staff members, it requires a lot of walking while measuring the resistivity of a large area. When measuring non-homogeneous soil, the resistivity obtained is the average of different layers of the soil involved. However, this result can be misleading if we intend to use the measured resistivity as an input to protect a metallic structure. For example, if we measure the surface resistivity to be of a certain value while the resistivity of the same soil a few depths deeper is way different than the surface resistivity, the material we intend to protect could be under or overprotected. Also, when testing the resistivity of a large area using Wenner's technique, it might become a tedious task to keep the separation of the four electrodes equally spaced, leading to unreliable results. Therefore, a better technique is needed to solve the challenges presented here.

3.1. Proposed Approach

The proposed method collects a dataset that is used as training and testing data to create a model that is utilized to predict the type of soil under investigation and aid engineers in selecting the recommended material that may withstand corrosion in that specific environment. Figure 3.1 shows the flowchart of the proposed solution. After collecting the dataset, a model is implemented and tested for prediction and measurement accuracy, precision, sensitivity, specificity, and hypothesis verification. An apparatus is built in the laboratory as a proof of concept to measure the resistivity of different types of soil. To address the issue caused by the continuous motion of the electrodes, five measurements

are taken. Four measurements are performed to form a square and the fifth sample is taken diagonally. The geometry used to measure electrical resistivity was defined by AEMC Instrument which is a leading distributor of tests, measurement, control, and calibration instrumentation. The geometric pattern setup is utilized in the AEMC 4630 Rechargeable Digital 4-Point Ground Resistance Tester to get a better estimation of soil resistivity of the grounding electrode site [14]. Since AEMC is a well-known company in the market and they have been successful using this measurement pattern (establishing it as an industry standard), it was utilized during the laboratory experiment. This proposed technique is expected to increase the accuracy when testing the resistivity of a non-homogeneous soil.

Unlike the device available in the market, the technique used in the device discussed in this dissertation will not just measure soil resistivity and help engineers to select the recommended material that may withstand corrosion in a given environment, but will also use a machine-learning algorithm to perform soil type prediction and dynamically improve its prediction performance. Upon completion, the proposed device will contribute to the advancement of the current state of the art technique used by the AEMC 4630 Rechargeable Digital 4-Point Ground Resistance Tester and the scientific literature research in artificial intelligence (AI), geology, agriculture, and civil engineering. The proposed device will leverage Wenner's four electrodes technique to extract parameters such as soil electrical resistivity, and moisture that may delay corrosion by selecting the recommended material that may withstand corrosion within a specific environment, and utilize the data collected over time to improve the machine learning prediction capability. Additionally, the device will contribute to the improvement of the current state of the art and may become one of the cutting-edge technologies used to perform subsurface geophysical surveys and metallic corrosion prevention.

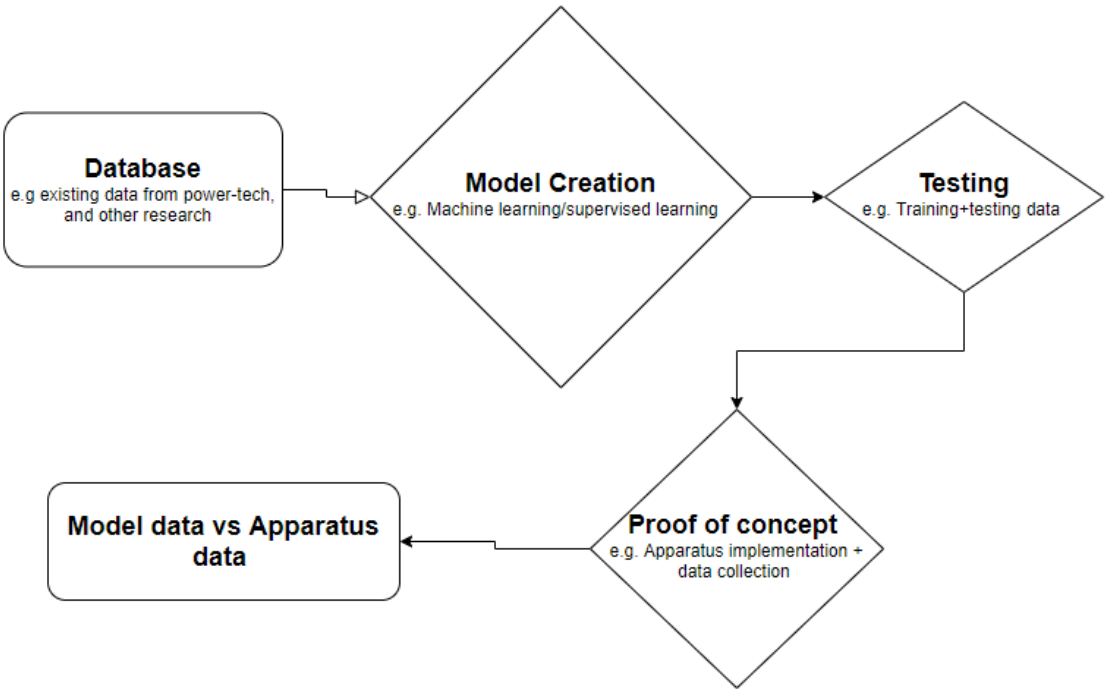

Figure 3.1. Project Flowchart

3.2. Prediction model

It is said that prevention is the best form of protection. The model presented here aims to simulate a hardware device that measures the electrical resistivity of soil. Then, from the electrical resistivity, the device is designed to predict the type of soil and estimate the corrosion aggressiveness of that soil. Once the aggressiveness of the soil is identified, the type of protection to protect the metal is defined. Several machine learning algorithms can be used to create this model. However, we have chosen a supervised learning approach to build the model, Figure 3.2. A supervised learning approach creates a model that makes predictions based on evidence in the presence of uncertainty [25]. As shown in Figure 3.2, this approach uses two techniques to create predictive models namely classification and regression. The supervised learning algorithm uses a set of predefined input and output data and trains a given model to generate reasonable predictions for the response to new data [25]. Within the supervised learning approach, several techniques could be chosen, but after careful analysis (trial and error of different algorithms), the Nearest Neighbor (k-NN) and Support Vector Machine (SVM) techniques were chosen

because they are excellent classifiers, they perform prediction by taking inputs, and they are easy to implement as opposed to, e.g., Artificial Neural Network (ANN), Decision Tree, Naïve Bayes, etc. The next section will provide more detail on the capability of each algorithm.

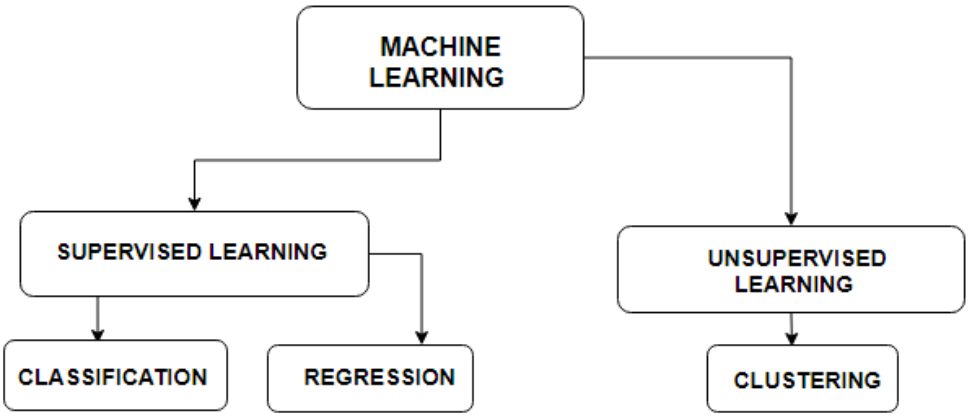

Figure 3.2. Machine learning techniques

3.3. Training the models - Nearest Neighbor (KNN) technique

The k-NN technique is a searching technique that locates all neighbors within a specified distance to query data points, based on the specified distance metric [26]. This classification technique allows the user to generate a search object with a training data set and passes the object and query data sets to the object functions. Additionally, the k-NN technique provides objects (KDTreeSearcher objects) that store the results (the training data, distance metric, and its parameters, and the maximum number of data points) of the nearest neighbor search that uses the kd-tree algorithm [27]. The kd-tree is a data structure used to split a space to organize points in a k-dimensional space. In this algorithm, the kd-tree is used to split the training data into two dimensions (x,y) to allow the KDTreeSearcher object to perform the searching of the k nearest neighbors of a testing point. Provided that a KDTreeSearcher object is created, the algorithm can find all neighboring points to the query data (testing data). To perform a nearest neighbor search, the KDTreeSearcher object is created first, then utilized to search the stored tree to find all neighboring points to the testing points stored within a data structure (e.g. Array) [28]. The algorithm (kd-tree) is more efficient than the exhaustive search algorithm (as

discussed in [29]) when the value of k (number of nearest neighbors to be found) is small ($k \leq 10$), the training and the testing data are sparse, and the training and query sets have many observations [28]. The KDTreeSearcher algorithm takes the training data with all its features as an input and partitions (fits) the data into regions according to the number of features available in the dataset. It is important to highlight that there is no single correct form to partition the data or plane into regions, different classification algorithms result in different partitions. Also, the value of k plays a pivotal role in the accuracy of the algorithm. The smaller the k value is the more likely the classification algorithm is to misclassify a testing point to belong to the wrong class. Therefore, it is a good practice to choose the value of k at least bigger than three to avoid misclassification due to outliers. Outliers are data points that lie outside its region or plane in the kd-tree partition, Figure 3.3. From figure 3.3, if we choose k = 1 for a given testing point that belongs to the black dots class, but it happens to be nearer one of the two circled blue dots class (outliers), the data would be misclassified as a blue dot class instead of black dot class. Therefore, if we select k to be at least bigger than three, we are less likely to misclassify our prediction.

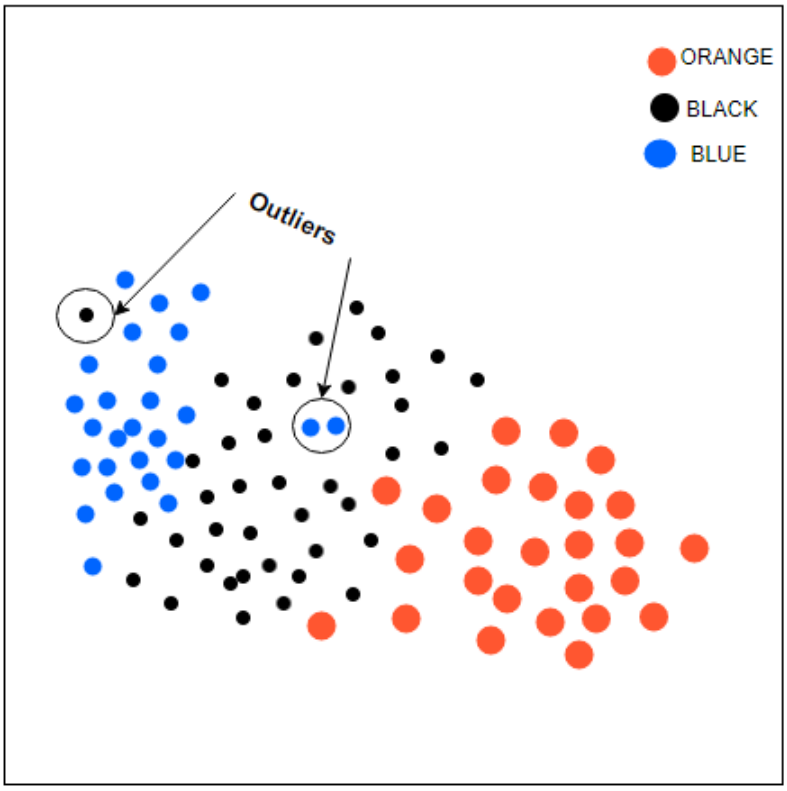

Figure 3.3. A plot displaying outliers

The k-NN function “knnsearch” takes the trained model, testing points, and the number of k-value as input. The function computes a search algorithm to return the indices of the closest points in the model for the input testing points. In addition to returning the indices of the closest points, the “knnsearch” function returns a matrix D which contains the distances between each observation for all k [30]. By using the indices of the closest points, the k training points closest to the n-th testing point is computed with high accuracy.

Since k-NN is capable of taking the soil resistivity, soil moisture, and soil labels as input, it became the most suitable technique to develop the model. Another reason is that the k-NN algorithm principle is very easy to implement. The k-NN algorithm only takes the input (training data), uses kd-tree data structure to organize the training data, and leverages kd-tree objects to find the k nearest points of the testing dataset. Therefore, if the k-NN algorithm is trained with a dataset containing the electrical resistivity and its moisture, the algorithm should be able to predict the type of soil and its corrosiveness for a new set of testing data. Finally, the ability of k-NN to display the input, output, and number of neighbors in the same graph is one of the reasons it was chosen.

3.4. KNN model results

The dataset in the appendix (Table A1) used in this model was collected from several scholars’ papers published in journals such as IEEE, Applied physics letters, Research gates, Elsevier, among others. These scholars conducted laboratory and field experiments to investigate the resistivity of different types of soils and the impact of the moisture content on its resistivity [3], [8], [9], [31], [32]. For the creation of the k-NN model, the collected dataset was divided into two sets: training (151 data points) and testing data (11 data points). The training data was used to create the model while the testing data was utilized for the proof of concept of the model. Both the training and testing data consists of a column of the electrical resistivity of different soils, the moisture content of the respective soils, detailed characterization of the type of soil, the site where the experiment was conducted or the soil sample was extracted, and lastly the depth of the measuring electrodes. The model was built using MATLAB (R2019b; The Mathworks)

which is a numerical computing software used to manipulate matrix, plotting data, implement models, create a user interface, etc.

In this approach, the following functions were used “KDTreeSearcher”, “knnseach”, and Tabulate. The model was first trained with a set of training and testing data. Then, a new dataset containing testing data never seen by the model was used to predict the type of soil. This approach is a classifier approach that allows the user to compute the k-nearest neighbor of each testing point value. Using the training dataset, a plot of the electrical resistivity in the function of the moisture content was obtained as shown in Figure 3.4. This figure shows pictorially the non-linear relation between electrical resistivity and moisture content. The different dots (or data points) plotted in figure 3.4 represents electrical resistivity and its corresponding moisture. As shown in figure 3.4 each colored dot corresponds to different types of soil. This figure also shows the electrical resistivity range of different types of soil as the moisture content is been changed.

Figure 3.4. A plot of the electrical resistivity against soil moisture

Figure 3.5 shows the plot of the testing data that verifies if the k-NN model estimates the type of soil when unknown resistivity and moisture values are input. The estimation of the type of soil is performed by the k-NN algorithm by calculating the k nearest neighbor of the input electrical resistivity and soil moisture. In other words, the k-NN looks for the closest electrical resistivity and soil moisture values to the input or training data. This estimation is done taking into consideration the training data so that the more training data the algorithm uses the more accurate the system becomes. The number of k-nearest neighbors is defined by the user. For example, we have set k to be equal to 4, so that the algorithm only identifies 4 types of soil from the predefined training points in which the input must belong. For example, figure 3.6 shows that the 4 nearest neighbors of the electrical resistivity with the value of 40.81 ohm*m (with 9.8% of water content) are: 38 ohm*m (16%), 34.37 ohm*m (13.5%), 35.5 ohm*m (15%), and 34.37 ohm*m (15.9%). Additionally, the model provides feedback in terms of the percentage of likelihood of a given electrical resistivity belonging to a specific type of soil. To illustrate, Figure 3.7 shows that the soil that measures an electrical resistivity value of 40.81 ohm*m (with 9.8% of water content) is more likely to be a sandy loam measured at a depth of 0.31m. As shown in Figure 3.7, the testing point has a 25% probability of belonging to either of the 4 nearest points. In a situation like this, the algorithm chooses the most likely among the four points, which is the first in the list in figure 3.7 (sandy loam). Figure 3.8 shows a different scenario for the training point of 14 ohm*m (49.01%). As illustrated, the measured electrical resistivity of 14 ohm*m (49.01%) is more likely to be a Clayey silt type of soil measured at a depth of 0.24m. As shown in [33] the data point of 14 ohm*m (49.01%) was extracted from a Clay type of soil. This shows that the k-NN algorithm was capable of predicting the clay type of soil based on its resistivity and moisture, but unable to detect the presence of silt in the soil. Table 3.1 summarizes the rest of the predictions for the testing points used to train the k-NN model.

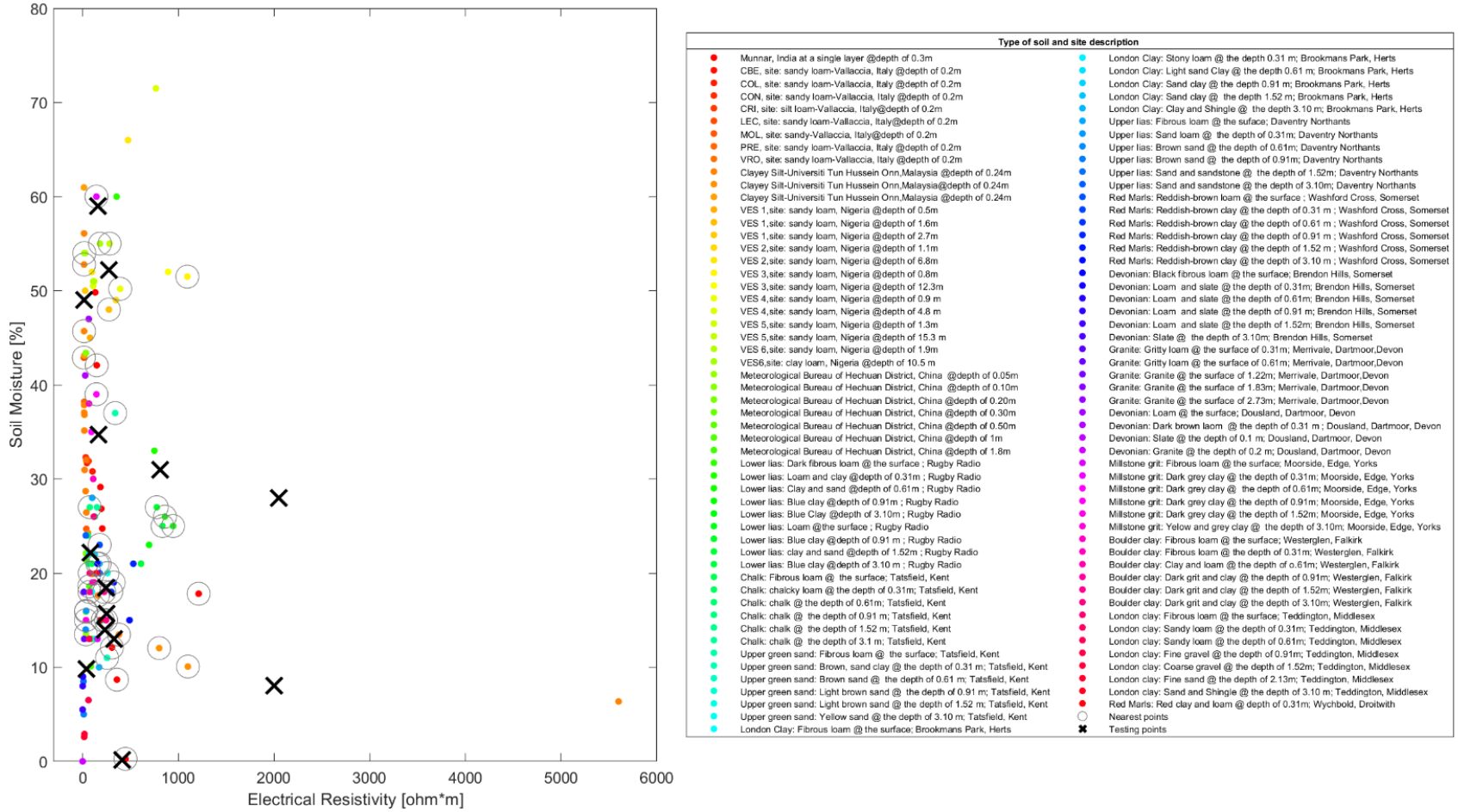

Figure 3.5. A plot of the testing points (x) showing the nearest neighbors (in a circle)

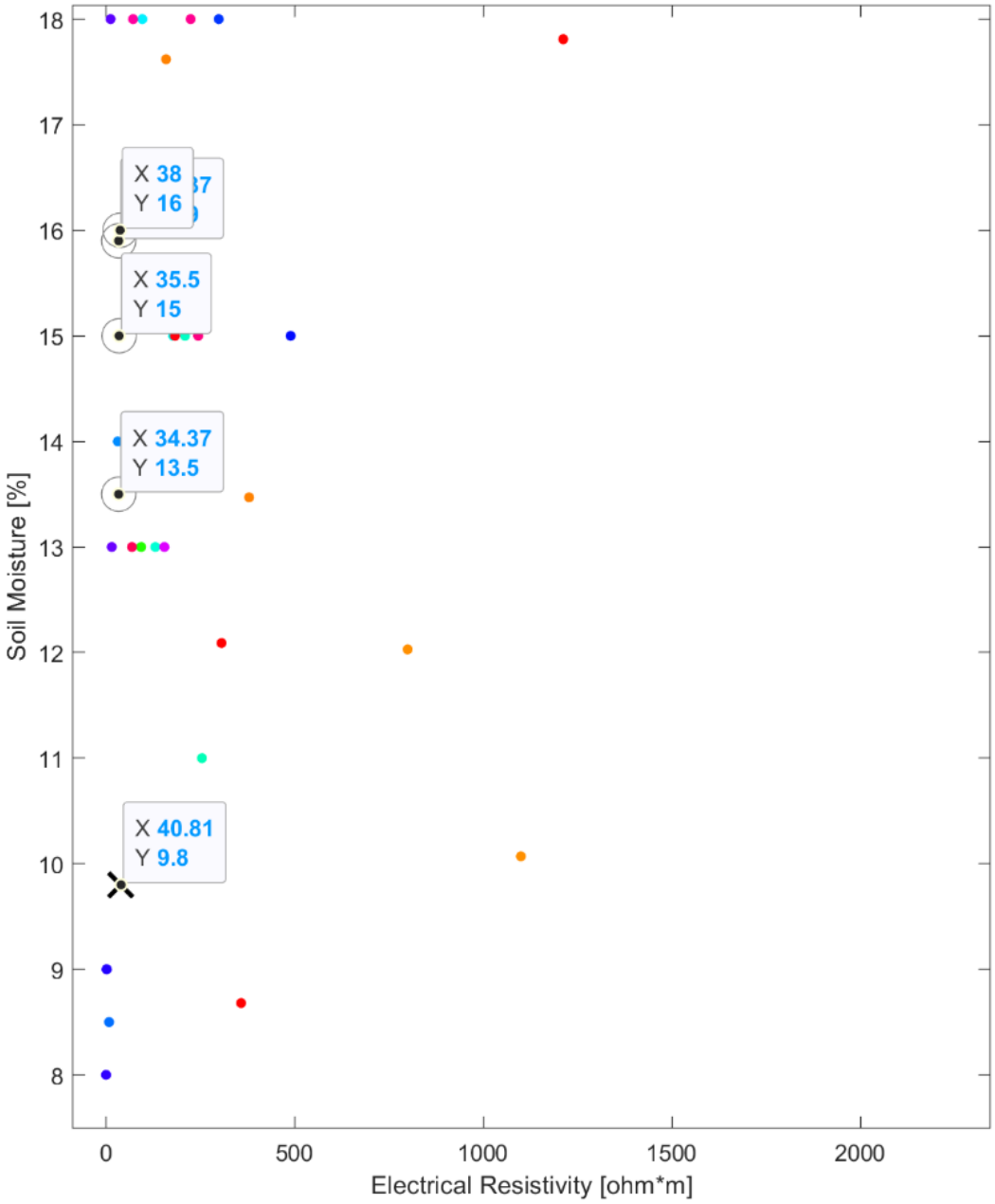

Figure 3.6. Four nearest neighbors for an electrical resistivity value of 40.81 ohm*m

Value	Count	Percent
Upper lias: Sand loam @ the depth of 0.31m; Daventry Northants	1	25.00%
Meteorological Bureau of Hechuan District, China @depth of 0.05m	1	25.00%
Millstone grit: Yelow and grey clay @ the depth of 3.10m; Moorside, Edge, Yorks	1	25.00%
Meteorological Bureau of Hechuan District, China @depth of 0.30m	1	25.00%

Figure 3.7. Probability of electrical resistivity to belong to a type of soil

Value	Count	Percent
Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m	2	50.00%
LEC, site: sandy loam-Vallaccia, Italy@depth of 0.2m	1	25.00%
VES6,site: clay loam, Nigeria @depth of 10.5 m	1	25.00%

Figure 3.8. Probability of electrical resistivity to belong to a type of soil

Table 3.1. Literature dataset prediction

Resistivity [ohm *m]	Corrosion level	Soil moisture	Depth (cm)	Actual soil:	Predicted
40.81	0	9.8	Unknown	Sandy Loam	Sandy Loam
14	8	49.01	Unknown	Clay	Clay Silt
340	0	37	Unknown	Fibrous Loam	Fibrous Loam
255	0	11	31	Sand Clay	Sand Clay
210	0	15	61	Brown Sand	Brown Sand
1	10	8	61	Loam and Slate	Loam and Slate
475	0	66	Unknown	Sandy Loam	Sandy Loam
250	0	15.67	Unknown	Clay Silt	Dark grit and clay
80	0	22.2	Unknown	Clay Silt	Sandy loam
160	0	59	Unknown	Sandy loam	Dark grey Clay
274	0	52.2	Unknown	Sandy loam	Sandy Loam

Table 3.1 summarizes the results obtained from the k-NN algorithms using 11 testing points. The model was built to predict the two laboratory soil (clayey silt and clay, soil 1 and 2 respectively). Therefore, the testing was conducted to identify clayey silt or clay (they are the true values). To understand the outcome and implications of the results presented in table 3.1, a confusion matrix was implemented. In statistical analysis, a confusion matrix is a table used to describe the performance of a classification model on a set of testing data for which the true values are known [34]. Figure 3.9 presents the confusion matrix of the results presented in Table 3.1 where,

- True Negative (TN) – Soils that are not clay/mixture of clay with a different soil predicted as not clay/mixture
- False Negative (FN) – Actual clay/mixture, but predicted as a different type of soil
- False Positive (FP) – A soil that is not actual clay/mixture, but was predicted as clay
- True Positive (TP) – Actual clay/mixture and predicted as clay/mixture
- n-Number of testing data

n = 11	Predicted: No	Predicted: Yes	
Actual: No	TN = 6	FP = 1	7
Actual: Yes	FN = 1	TP = 3	4
	7	4	

Figure 3.9. Literature/Training Data Confusion Matrix

Accuracy = $\frac{(TP+TN)}{n} = \frac{9}{11} = 0.82$ eq. 3.1

Error Rate = $\frac{FP+FN}{n} = \frac{2}{11} = 0.18$ eq. 3.2

Sensitivity= $\frac{TP}{Actual{:}Yes} = \frac{3}{4} = 0.75$ eq. 3.3

Specificity = $\frac{TN}{Actual{:}\,No} = \frac{6}{7} = 0.86$ eq. 3.4

Precision = $\frac{TP}{Predicted{:}Yes} = \frac{3}{4} = 0.75$ eq. 3.5

It is important to mention that the prediction was conducted considering clayey silt and clay as the true positive (TP). Based on the 151 points used to train the model, the eleven testing points were predicted with accuracy, error rate, sensitivity, specificity, and precision of 82%, 18%, 75%, 86%, and 75%. These results show that the testing data fit the model very well with only a 0.18 error rate. The high sensitivity value (75%) indicates the model is capable of predicting the true name of an unknown soil 75% of the

time. On the other hand, the specificity value (86%), implies that the algorithm or model implemented here is capable of predicting when certain soil is not clayey silt or clay 86 % of the time. In section 4.3.2 (Data analysis), the model will be tested with the data collected in the laboratory and compared with the literature model results presented in this section.

Once the type of soil for a given electrical resistivity value is identified, the model can estimate the corrosiveness of the soil. The AWWA standard uses a scale from 1-10 wherein 10 indicates soil corrosiveness to ductile-iron pipe [15]. The training and testing data resistivity values are mapped to the AWWA standard scale to determine its corrosion level to the ductile-iron pipe. One of the missions of the AWWA is "to review interior and exterior corrosion of ductile iron pipe and fittings and to draw standards for the interior and exterior protection of ductile-iron pipe and fittings" [15]. It is important to mention that the AWWA corrosion characterization, Table 3.2, discussed in this dissertation was utilized for illustration purposes only. In other words, given that the resistivity of soil and the type of soil under investigation is known, the corrosion level of a given soil is determined, thus the best material can be selected by taking into consideration the ability of the material to withstand corrosion in that environment. However, one should not use the AWWA corrosion level classification discussed here to make conclusive decisions about a real-world project until field testing is conducted, and all the limiting variables are considered. Table 3.2 shows a table extracted from [15] where the resistivity range and corrosion level for a ductile-iron pipe are presented. In the case of an electrical resistivity value of 14 ohms*m, the level of corrosion is 10, and protection is required if a ductile-ion pipe is to be placed in the ground.

Table 3.2. Soil test evaluation (adapted from [15])

Soil Characteristics Based on Samples Taken Down to Pipe Depth		
Resistivity – ohm-cm (based on water-saturated soil box):		*Points
Resistivity	< 1500 (15 ohm*m)	10
	>= 1500 - 1800	8
	>1800-2100	5
	>2100 - 2500	2
	>2500 - 3000	1
	>3000	0

*Ten points or greater indicates that soil is corrosive to the ductile-iron pipe; protection is needed.

3.5. Support Vector Machine

The SVM is another supervised learning classification algorithm used to separate two sets of data. This algorithm works by finding the best hyperplane that separates all data points of one class from those of the other class [35]. The optimal hyperplane for an SVM technique is the one with the largest margin (maximal width of the slab parallel to the hyperplane that has no interior data points, Figure 3.10) between two classes [35]. Within the SVM algorithm, there are several fitting functions used to fit the data. However, this dissertation used “fitcecoc” [35] algorithm to training and separate five classes of soil namely sandy loam, clayey silt, unknown soil 1, unknown soil 2, and blue clay. The error-correcting output codes (ECOC) is a classifier “for multiclass learning, where the classifier consists of multiple binary learners such as support vector machines (SVMs)” [36].

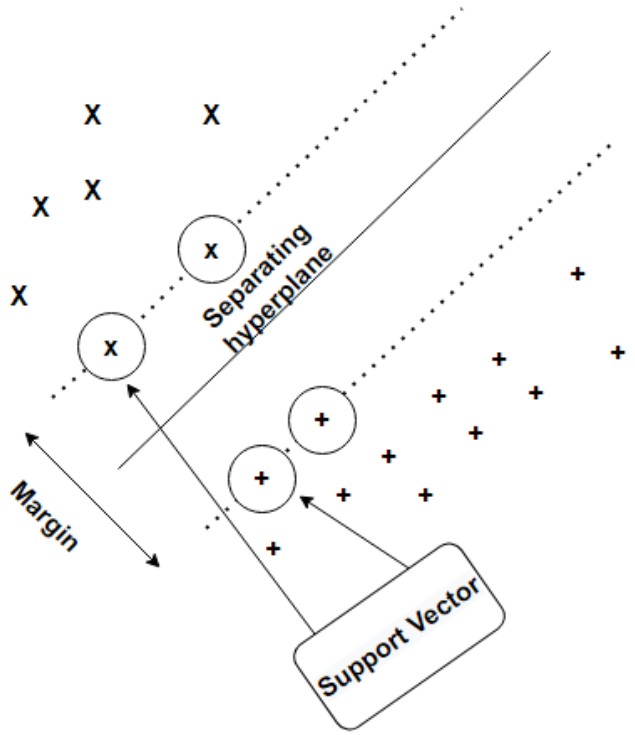

Figure 3.10. A plot indicating support vector machine parameters

3.6. SVM model results

The data used in this model was collected from the same scholars' papers as in the previous model (see table A2 in the appendix). However, only five classes of the collected data were used to simplify the algorithm and the interpretation of the collected data. Figure 3.11 shows the plot of the five classes of data in small dots of different colors. The algorithm learned the data and the SVM technique was applied to separate the five classes. In this specific model, five learners corresponding to each of the five soil types were created. The algorithm performed several iterations (one-versus-one coding) wherein the first SVM binary leaner performed all observations and classifications using sandy loam in red dot (figure 3.11). The circles shown in figure 3.11 indicate the support vectors for each of the classes used in the model. The data enclosed by either circle are predicted to be part of the support vector class enclosing the data. To illustrate, the data points 18.56 (90% moisture) are defined as unknown soil 1 (in green dot) in figure 3.11, however, it was enclosed by SVM 2 (clayey silt) which means that the unknown soil is clayey silt. The error of the model was calculated to be 44.07 % which is quite large.

This large percentage of error was evident in the performance of the algorithm. For example, several blue clay soils (892(5% moisture), 1094 (51.5% moisture), 695 (23% moisture), etc) were classified to be SVM 4 (unknown soil 2). We know that these classifications are incorrect because the unknown soil 2 (or clay) used in the laboratory experiment is not blue clay as predicted by the model. Therefore, the SVM is not a good model to predict the type of soil based on the resistivity of soil and its moisture. Another reason for the inaccuracy of the SVM algorithm might be that the training data used to create the model cannot be easily separated by the algorithm into different classes, resulting in a high error rate of 44.1%.

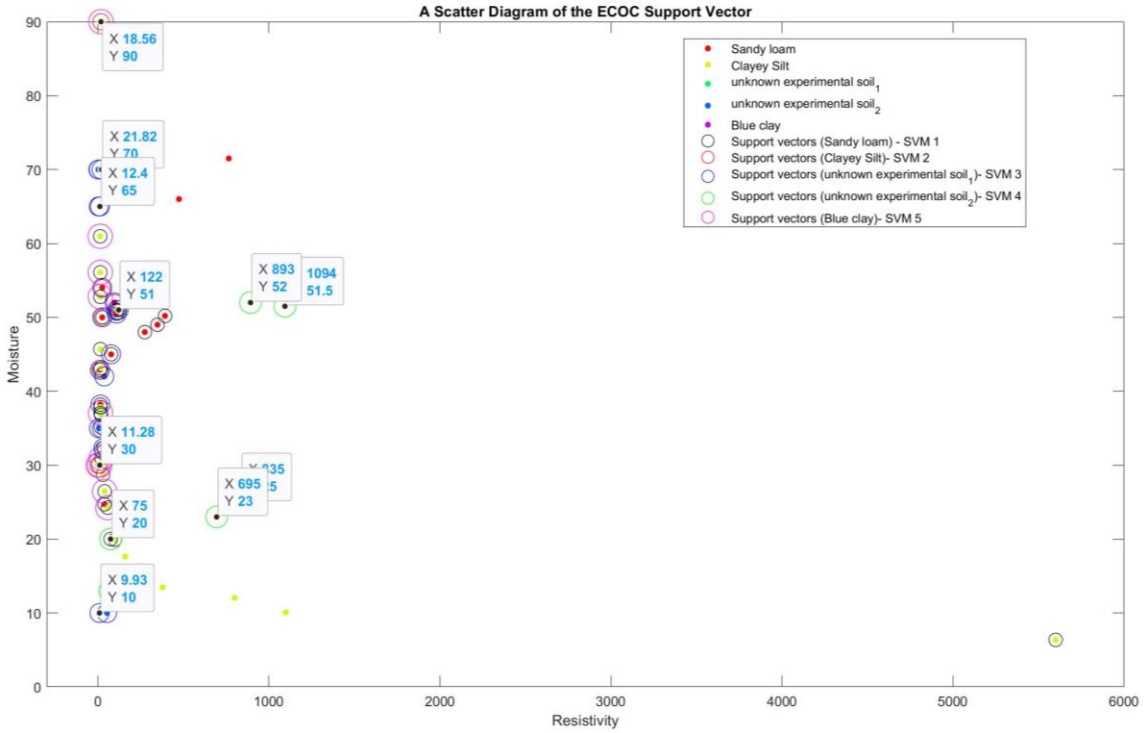

Figure 3.11. A plot displaying support vectors

Chapter 4.

Preliminary experimental verification

The results obtained from the model were intended as a proof of concept that the proposed approach may be a viable solution to help prevent corrosion, through proper materials selection, on underground metallic structures. However, a laboratory experiment was next conducted to further examine the hypothesis and machine learning model. The data collected during the laboratory experiment is used to compare with the data obtained from the k-NN model created on MATLAB. The next section of this thesis introduces the techniques used during the laboratory experiment and discusses in detail the collected data and its implications.

4.1. Experimental Setup

The experiment was conducted within a room in which the temperature varied from 24-26 ^{o}C. For each measurement, the temperature was taken using a temperature sensor attached to a digital multimeter (DMM) as shown in figure 4.1. The voltage drop across the two inner electrodes was measured using the DMM as shown in figure 4.2. The DMM was reset after every measurement to avoid any measurement error that might be introduced. On the other hand, the humidity of the soil was measured using the "3-way soil meter" (figure 4.3). This meter is commonly used to measure garden moisture, pH, and the intensity of light. The accuracy of the "3-way soil meter "is not the main goal of this research. To ensure that the soil moisture is being properly measured, several measurements across the soil sample was taken and an average of the measured soil moisture was calculated.

Two different types of soil were used in this experiment. One of the soil was homogeneous (same type of soil) while the other was a mixture of two different forms of soil. Both soils were labeled as unknown (unknown soil 1 and 2) because no description of the soil was provided by the seller. After consulting an Oil Gas & Salt Resources Library geologist employee, soil 1 was identified as clayey silt and soil 2 as clay (A.

Cachunjua, personal communication, August 30th, 2020). However, during the experiment, both soils 1 and 2 are considered as unknown to be predicted by the algorithm. The fact that both soils were unknown is a good thing because we intend to estimate their characteristic by using the k-NN algorithm. Soil 1 was placed in a (31.2cm) x (17.5cm) x (11.3cm) box, length, width, and height, respectively. Figure 4.4 shows the box in which the experiment for soil 1 was conducted. For this soil, only two sets of experiments were conducted specifically for electrodes of depth 1cm (2cm separation) and 2cm (4cm separation). However, soil 2 was placed in the box shown in figure 4.5 which possesses the following dimensions: (37.1cm) x (29cm) x (17.2cm). Three experiments were conducted using soil 2 (box 2-figure 4.5) because it is slightly bigger than box 1 used for soil 1. The experiments conducted on box 2 (soil 2) were for electrodes of depth 1cm (2cm separation), 2cm (4cm separation), and 4cm (8cm separation).

Figure 4.1. DMM and temperature sensor

Figure 4.2. A DMM used to measure the voltage drop across two inner electrodes

Figure 4.3. 3-way soil meter

Figure 4.4. Box 1 a (31.2cm) x (17.5cm) x (11.3cm) and soil 1

Figure 4.5. Box 2 (37.1cm) x (29cm) x (17.2cm) and soil 2

Since four electrodes are needed in Wenner's technique for soil's electrical resistivity measurement, four electrodes were also used in this experiment (figure 4.6). As suggested by [4] the electrodes are fabricated of stainless steel. The diameters of the electrodes are 1cm and the total length of each electrode is approximately 10.5cm long. To keep track of the depth of penetration of each electrode in the soil, white marks were placed at 1cm, 2cm, and 4cm depth along the electrode's length.

4.2. Electrode setup technique

In each experiment, the electrical resistivity of soil 1 and 2 was measured at three different depths specifically at 1cm, 2cm, and 4 cm of electrode penetration. The electrodes were separated at different distances for each of the depths. Figure 4.7 shows the electrode placement schematics at three different depths. For example, as illustrated in figure 4.7, at the depth of 4cm the electrodes are equally separated by 8cm or double the depth of penetration. To help to keep the electrodes equally spaced a placeholder was designed for each experiment (1cm, 2cm, and 4cm depth), figure 4.8.

Figure 4.6. Stainless steel electrodes

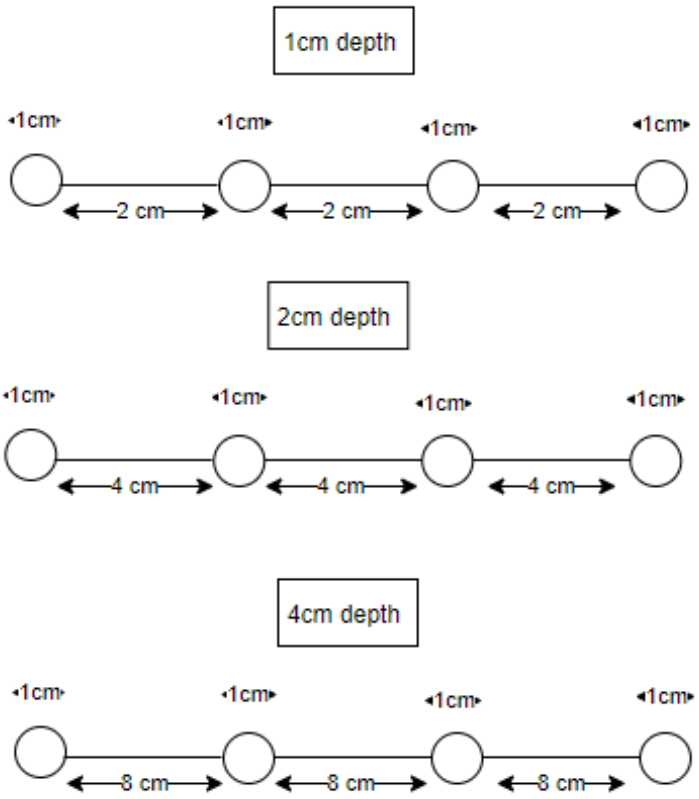

Figure 4.7. Electrodes placement setup

Figure 4.8. Electrode separation placeholder

To improve the accuracy of the measurement, five voltage drop samples were taken for each of the soil moisture percentages. Figure 4.9 shows each of the five measurement positions for each of the soil moisture percentages. As shown in figure 4.9, measurements 1 to 4 were taken parallel to each other while sample five was taken diagonally. The voltage drop of each measurement was added and divide by 5 to find the voltage drop of that specific area and soil moisture value eq.4.1. Additionally, several voltages drop measurements were performed for a single moisture value to account for the shaking electrodes which can introduce measurement errors and voltage reading inaccuracy. The electrical resistivity of the soil was then calculated using equations eq 4.2 and 4.3. Tables 4.1- 4.5 below show the data collected in each of the measurements for each of the experimental depths (1cm, 2cm, and 4cm of electrode penetration).

$$voltage\ drop\ avg = \frac{vm1+vm2+vm3+vm4+vm5}{5} \qquad \textbf{eq.4.1}$$

$$R = \frac{voltage\ drop\ avg}{Iavg} \qquad \textbf{eq.4.2}$$

$$\rho, \Omega \cdot \mathrm{m} = 2\pi a * \mathrm{R}\ (\text{a is in meters}) \qquad \textbf{eq.4.3}$$

Where, *vm1,2,3,4,5* are the voltage drop at each sample in volts (V)

Iavg, input current in Ampere (A)

a, electrodes separation in meters (m)

ρ, soil electrical resistivity in $\Omega \cdot \mathrm{m}$

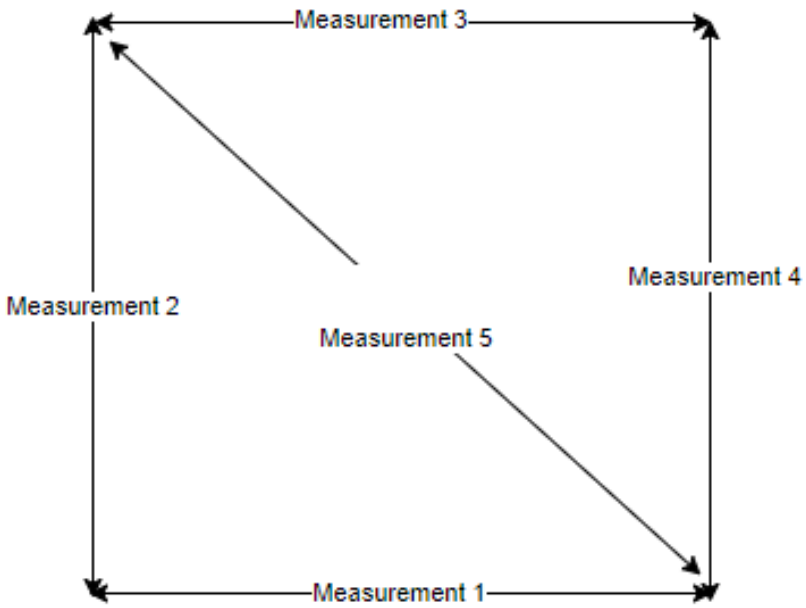

Figure 4.9. Voltage drop measurement sample

Table 4.1. Soil 1-at 1 cm depth

Voltage drop trials (V)					Current (A)					R (Ω)	a (m)	Tempe-rature	Soil Moisture	P (Ω · m)
0.26	0.25	0.3	0.52	0.17	0.02	0.02	0.02	0.02	0.02	15	0.02	24	30	1.88
2.28	3.12	3.25	2.4	2.2	0.04	0.04	0.04	0.04	0.04	66.25	0.02	24	90	8.33

Table 4.2. Soil 1-at 2 cm depth

Voltage drop trials (V)					Current (A)					R (Ω)	a (m)	Tempe-rature	Soil Moisture	P (Ω · m)
1.3	1.22	1.22	2	0.99	0.03	0.03	0.03	0.03	0.03	44.87	0.04	24	30	11.28
3.66	3.2	4.8	2.8	4	0.05	0.05	0.05	0.05	0.05	73.84	0.04	24	90	18.56

Table 4.3. Soil 2-at 1 cm depth

Voltage drop trials (V)					Current (A)					R (Ω)	a (m)	Tempe-rature	Soil Moisture	P (Ω · m)
0.13	0.2	0.3	0.45	0.35	0.02	0.02	0.02	0.02	0.02	14.30	0.02	24	10	1.80
0.7	1.82	1.65	1.2	1.6	0.02	0.03	0.03	0.02	0.03	52.80	0.02	25	35	6.64
1.3	1.6	1.2	1.2	1.2	0.03	0.03	0.03	0.03	0.03	43.33	0.02 0.02	24	65	5.45
0.5	0.75	1.2	0.8	0.7	0.03	0.03	0.03	0.03	0.03	26.33	0.02	24	70	3.31

Table 4.4. Soil 2-at 2 cm depth

Voltage drop trials (V)					Current (A)					R (Ω)	a (m)	Tempe-rature	Soil Moisture	P (Ω · m)
1.2	0.45	0.6	0.7	1	0.02	0.02	0.02	0.02	0.02	39.50	0.04	24	10	9.93
3	3.3	2.8	2.3	2.5	0.03	0.03	0.03	0.03	0.03	92.67	0.04	25	35	23.29
1.6	1.8	1.7	1.3	1	0.03	0.03	0.03	0.03	0.03	49.33	0.04	24	65	12.40
0.8	0.6	0.3	0.7	0.8	0.03	0.03	0.03	0.03	0.03	21.33	0.04	24	70	5.36

Table 4.5. Soil 2-at 4 cm depth

Voltage drop trials (V)					Current (A)					R (Ω)	a (m)	Tempe-rature	Soil Moisture	P (Ω · m)
2.2	2.24	2.2	2	2.1	0.02	0.02	0.02	0.02	0.02	107.4	0.08	24	10	53.99
3.2	3.3	2.4	3.5	3	0.03	0.03	0.03	0.03	0.03	102.67	0.08	26	35	51.61
2.3	2.5	2.55	1.7	2.3	0.03	0.03	0.03	0.03	0.03	75.67	0.08	24	42	38.03
1.3	1.2	1.3	1.31	1.4	0.03	0.03	0.03	0.03	0.03	43.40	0.08	24	70	21.82

4.3. Results and Discussion

Figure 4.10 shows the final laboratory experimental setup where a current is applied at the outer electrodes and the voltage drop is measured from the two inner probes. A Direct Current (DC) power supply was used. The current density increases within conductive regions and decreases within resistive regions. The depth of penetration of the input current or electrical signal depends on the instrument's strength of the signal [6]. The electrical charges from the current input build-up at the interfaces between regions of different electrical conductivity [23]. These variations in charge are then detected by the two inner electrodes and registered as variations in the distribution of potential (voltage drop). After performing several measurements, the collected data was input into an excel spreadsheet to compute all the necessary magnitudes. A MATLAB file was then created from the excel spreadsheet to allow the algorithm to interpret the collected data.

Figure 4.10. Wenner's four-electrode box experiment

4.3.1. Mean value and standard deviation of the experimental data

Since five measurements were taken to determine the electrical resistivity for a given depth and soil moisture (figure 4.9), there were some differences in the measured voltages, currents, and depth for each sample (tables 4.1-4.5 and 6.3-6.7). Therefore, the mean value and standard deviation for each measurement were calculated. Tables 4.6-4.10 show the mean value and standard deviation for each soil and different depths. As shown below, the mean voltage value and standard deviation are all within 1 standard deviation. It means that each one of the five voltages drops measured for each sample was very close to the mean value. The same can be said for the current mean value and standard deviation. As shown in the tables below, the current standard deviation for most of the samples was calculated to be zero. These results were expected since the current was constant for most of the experimental samples. The equations below were used to calculate the mean and standard deviation from tables 4.1-4.5.

$$\text{Mean} = \frac{1}{\mathrm{N}} * \sum_{i=1}^{N} mi \quad \textbf{eq.4.4}$$

$$\text{Standard Deviation} = \sqrt{\left(\frac{1}{N} \sum_{i=1}^{N} (m_i - Mean)^2\right)} \quad \textbf{eq.4.5}$$

In equations 4.4 and 4.5, m is the measurement while i and N are the sample number and the total number of measurements.

Table 4.6. Mean value and standard deviation of soil 1-at 1 cm depth

Voltage Mean (V)	Voltage Standard Deviation (V)	Current Mean (A)	Current Standard Deviation (A)
0.3000	0.1317	0.0200	0
2.65	0.4957	0.0400	0

Table 4.7. Mean value and standard deviation of soil 1 at 2 cm depth

Voltage Mean (V)	Voltage Standard Deviation (V)	Current Mean (V)	Current Standard Deviation (V)
1.62	0.2490	0.0300	0
3.10	0	0.0520	0.0084

Table 4.8. Mean value and standard deviation of soil 2 at 1 cm depth

Voltage Mean (V)	Voltage Standard Deviation (V)	Current Mea (A)n	Current Standard Deviation (A)
0.2860	0.1254	0.0200	0
1.3940	0.4496	0.0200	0.0055
1.30	0.1732	0.0300	0
0.7990	0.2559	0.0300	0

Table 4.9. Mean value and standard deviation of soil 2 at 2 cm depth

Voltage Mean (V)	Voltage Standard Deviation (V)	Current Mean (A)	Current Standard Deviation (A)
0.7900	0.3050	0.0200	0
2.7800	0.3962	0.0300	0
1.4800	0.3271	0.0300	0
0.6400	0.2074	0.0300	0

Table 4.10. Mean value and standard deviation of soil 2 at 4 cm depth

Voltage Mean (V)	Voltage Standard Deviation (V)	Current Mean (A)	Current Standard Deviation (A)
2.1480	0.0976	0.0200	0
3.0800	0.4207	0.0300	0
2.2700	0.3384	0.0300	0
1.3020	0.0709	0.0300	0

Two measurements at the same depth and soil moisture were taken to calculate sample errors under the same conditions. Measurement at the depth of 2 cm, at the moisture of 90 %, and temperature of 27 ^{o}C was taken and compare to a previous measurement at the same condition, but different temperature (at 24 ^{o}C). The first measurement taken from soil 1 at the temperature of 24 ^{o}C, 2 cm depth, 4cm separation, and 90 % moisture had an electrical resistivity of approximately 18.56 ohm*m while the second measurement performed at the same condition, but at a temperature of 27 ^{o}C had an electrical resistivity around 15.32 ohm*m. It becomes clear that there is a small discrepancy between the two electrical resistivities (18.56 ohm*m (at 24 ^{o}C) and 15.32 ohm*m (27 ^{o}C)). By subtracting the first and second electrical resistivity, the discrepancy is calculated to be 3.24 ohm*m. Assuming a normal distribution, the mean value of the two electrical resistivities is calculated to be about 16.94 ohm*m while the standard deviation becomes 2.229 ohm*m. Taking into consideration that the measurement considered herein were performed at different temperatures (about 3 ^{o}C of difference), the discrepancy of the

two measurements could be considered as small enough. This implies that a small change in temperature does not cause a dramatic change in the resistivity of a given soil. Also, the small standard deviation (about 2.229 ohm*m) implies that the two measured electrical resistivities are within 1 standard deviation. In other words, the Wenner technique used to measure soil resistivity during this experiment is accurate enough and the electrical resistivity of the soil does not change drastically with a small variation in temperature.

4.3.2. Data analysis

In this section, the k-NN model created in section 3.3 is tested using the data collected from the laboratory experiment and a comparison between the literature and laboratory data performance is conducted. Table 4.11 below presents the laboratory experimental data used to test the model to predict the laboratory soil samples. To verify the performance of the model to the experimental data, ten samples collected from the lab were input into the model specifically 21.82 ohm*m [70%], 38.03 ohm*m [42%], 12.40 ohm*m [65%], 9.93 ohm*m [10%], 6.64 ohm*m [35%], 18.56 ohm*m [90%], 8.34 ohm*m [90%], 1.80 ohm*m [10%], 5.45 ohm*m [65%], and 53.99 ohm*m [10%]. These datasets were never seen by the model, and they were extracted from the laboratory soil samples (clayey and clay, soil 1 and 2, respectively). As shown in table 4.11 six sample points (in green) collected from the laboratory were correctly predicted as clayey silt and clay. This result shows that the model is capable of predicting the type of soil given that the resistivity and moisture of the soil are known. This result also proves the hypothesis of this thesis that a k-NN machine learning algorithm is capable of predicting the type of soil of an unknown soil if the resistivity and moisture of that soil are known.

Table 4.11. Experimental dataset prediction

Resistivity [ohm *m]	Corrosion level	Soil moisture	Depth (cm)	Actual soil:	Predicted
21.82	2	70	4	2(clay)	Clay Loam
38.03	0	42	4	2 (clay)	Sandy Loam
12.40	10	65	2	2 (clay)	Clayey Silt
9.93	10	10	2	2 (clay)	Brown Sand
6.64	10	35	1	2 (clay)	Clayey Silt
18.56	5	90	2	1 (clay)	Clayey Silt
8.34	10	90	1	1 (clayey silt)	Clayey Silt
1.80	10	10	1	1 (clayey silt)	Loam
5.45	10	65	1	2 (clay)	Clayey Silt
53.99	0	10	4	2 (clay)	Chalk Loam
14	8	49.01	Unknown	Clay	Clay Silt
340	0	37	Unknown	Fibrous Loam	Fibrous Loam
255	0	11	31	Sand Clay	Sand Clay
210	0	15	61	Brown Sand	Brown Sand
1	10	8	61	Loam and Slate	Loam and Slate
475	0	66	Unknown	Sandy Loam	Sandy Loam
250	0	15.67	Unknown	Clay Silt	Dark grit and clay
80	0	22.2	Unknown	Clay Silt	Sandy loam
160	0	59	Unknown	Sandy loam	Dark grey Clay
274	0	52.2	Unknown	Sandy loam	Sandy Loam

The overall performance of the model was analyzed using a confusion matrix as performed on the literature sample in chapter 3, but a new dataset collected from the laboratory experiment was added to the testing data. Figure 4.11 below displays the confusion matrix of the experimental data. Using a total of 20 points the confusion matrix was created wherein:

- True Negative (TN) – Soils that are not clay/mixture of clay with a different soil predicted as not clay/mixture

- False Negative (FN) – Actual clay/mixture, but predicted as a different type of soil
- False Positive (FP) – A soil that is not actual clay/mixture, but was predicted as clay
- True Positive (TP) – Actual clay/mixture and predicted as clay/mixture
- n-Number of testing data

n = 20	Predicted: No	Predicted: Yes	
Actual: No	TN = 5	FP = 1	6
Actual: Yes	FN = 5	TP = 9	14
	10	10	

Figure 4.11. Laboratory/Experiment Sample Data Confusion Matrix

Accuracy = $\frac{(TP+TN)}{n} = \frac{14}{20} = 0.70$ eq.4.6

Error Rate = $\frac{FP+FN}{n} = \frac{6}{20} = 0.3$ eq.4.7

Sensitivity= $\frac{TP}{Actual:Yes} = \frac{9}{14} = 0.64$ eq.4.8

Specificity = $\frac{TN}{Actual: No} = \frac{5}{6} = 0.833$ eq.4.8

Precision = $\frac{TP}{Predicted:Yes} = \frac{9}{14} = 0.64$ eq.4.10

In terms of accuracy, the model performed well with 70% accuracy, which indicates that the model is capable of predicting the correct type of soil 70 percent of the time. This accuracy percentage is similar to the accuracy obtained by Lemercier in [12] where a boosted classification tree was used to predict soil parent material and drainage. An accuracy of about 70% was also described as adequate in [12] and [13]. The accuracy achieved by the k-NN model created here is good enough compared to [12] and [13] and considering the fact that the data used to train the model were extracted from scholarly sources in which the accuracy and veracity of the data are unknown. On the other hand, the error rate of the model is 0.3 (or 30%) which means that the model is capable of predicting the right type of soil 70 percent of the time. This error rate can be decreased by increasing the number of training sample points. In contrast, the sensitivity of the model is 64% percent which indicates that we can only be sure of a true positive (TP) 64% of the time. This sensitivity result is a bit low because the higher the sensitivity value the better the model becomes. However, the specificity result is 83.3 % which indicates that we can be sure that a true negative (TN) value is indeed a correct prediction 83.3 percent of the time. In terms of precision, the model proved to be very precise. The precision value of 90% indicates that the model is capable of precisely predict the correct type of soil 90 percent of the time.

Table 4.12. Comparison between literature and experimental dataset models

	Statistical Results Comparison	
	Literature Dataset Model	Experimental Sample Model
Accuracy	0.82	0.70
Error Rate	0.18	0.3
Sensitivity	0.75	0.64
Specificity	0.86	0.833
Precision	0.75	0.64

Table 4.12 above shows the comparison between the statistical results obtained using the literature testing dataset (used to test the k-NN model in chapter 3) and the experimental dataset (used to predict the laboratory soil samples). As shown the literature testing data (accuracy, error rate, sensitivity, specificity, and precision equal to 82%, 18%, 75%, 86%, and 75% respectively) fits better the model than the experimental testing data points (accuracy, error rate, sensitivity, specificity, and precision equal to 70%, 30%, 64%,

83.3%, and 64% respectively). The reason for the literature data points fitting better the model might be because most of the training data were collected from the literature papers. The model might be biased to the dataset extracted from the scholars' literature, thus the accuracy, error rate, sensitivity, specificity, and precision are better than the ones obtained from the experimental dataset.

In terms of sample size, the model would perform better if the training dataset was bigger than 151 sample points (6 points were added later from the laboratory measurement). The model would have more points to train the algorithm and the class of electrical resistivity for different types of soil would be more evident or easier to be classified by the algorithm. However, if the training data’s sample size becomes too big, it might lead to overfitting which might impact the performance of the algorithm negatively. Overfitting happens when the model learns all the characteristics of the training data to the point of negatively impact the performance of the model. Therefore, careful consideration is required when increasing the number of datasets used to create a model.

Chapter 5.

Limitations

The results presented in the last section of this dissertation provides initial proof-of-concept verification of the hypothesis of this research. However, those results were influenced by several factors that must be reported. The main limitation of this research is the resistivity meter accuracy, depth of measurement, electrode separation, the temperature at which the samples were taken, origin of the data used for model creation, experiment site, soil moisture, and model's bias.

As previously stated, the data used to create the machine learning model was collected from different scholars' papers. This research assumes that the collected data are accurate, and the resistivity meter used by these scholars was also accurate. Also, it was assumed that these scholars did not make any computation mistakes related to the reported data. On the other hand, it was shown that the resistivity of soil varies according to the depth and separation of the probe's electrodes. This report assumes that the data reported by each scholar was extracted taking into consideration these parameters and no measurement mistakes were made.

Some of the data used in this report were extracted from scientific papers that did not provide enough detail about the temperature at which each measurement was collected. It was found that during the laboratory experiment that the temperature in which the experiment is conducted affects the resistivity of soil. Thus, it would be important to know the temperature at which the experiments were conducted to better train the machine learning models.

The homogeneity of the dataset used is another limitation of this research since the data were taken from different literature. There are no standards to determine how homogeneous the data from different literature are to each other. Therefore, the data collected in the laboratory were compared against the literature data to help to decide which data from the literature were reasonable to create the model. The parameters used

to decide either a given literature data should be included are temperature and electrical resistivity. From the laboratory experiment, It was observed that the electrical resistivity increased as the temperature increased. Therefore, the datasets of literature with the same characteristics as the laboratory measurements were included to create the model. Additionally, the datasets from different literature were added to the training model separately and the accuracy of the model was calculated afterward. Datasets that decreased the accuracy of the model by 40% were excluded.

The percentage of water content in the soil was varied from oven-dry (approximately 0%) to about 90% water content for each depth tested in the laboratory experiment. However, after pouring 660ml of water for each depth of the measurement sample, the soil was not evenly wet. This resulted in different soil moisture percentage over the surface area of the soil even after the soil was properly mixed. Therefore, the average of different soil moisture measurement was taken and used as the moisture level. This unequal moisture percentage across the soil for a given depth sample might negatively influence the results reported from the laboratory experiment. In addition, soil resistivity, even for a specific soil type (e.g., clay) may be highly dependent on depth in the field as characteristics such as moisture and density change with depth.

Finally, the data used to implement the model were divided into training and testing data. In some cases, the model performed well with the training data but performed poorly when unseen data was input as the testing data. This machine learning problem is known as overfitting (the opposite of this issue is called underfitting). Although the k-NN and SVM models were implemented taking into consideration these issues, it is possible that the results of the model were biased by overfitting or underfitting problems.

5.1. Model limitation

The accuracy of the algorithms is biased by the parameters used as training data. In other words, the more features the algorithm uses to predict the unknown variable, the more effective the algorithm becomes. However, in this research, only two main features were used to verify the hypothesis namely soil resistivity and moisture. In reality, soil moisture

and resistivity are not the only factors that should be used to predict the type of soil. As found out during the experiment temperature variation, and electrodes depth of penetration are relevant parameters that affect the resistivity of soil, thus these parameters should be considered. The model would predict better the type of soil when only the electrical resistivity of the soil is known if electrodes depth of penetration and temperature were also used to train the model.

Additionally, the pH of the soil is another factor that should be considered when selecting the recommended material that may withstand corrosion within a given environment. The pH of a given soil affects the corrosion of metallic structures such as pipes underground [37]. Therefore, not using the pH of the soil as a parameter to predict different types of soil and selecting the recommended material that may withstand corrosion in a specific environment is another limitation of the model implemented in this dissertation.

In short, the limitations reported above might have influenced the results reported in this research. However, these limitations did not affect the proof-of-concept effectiveness of the proposed machine learning algorithm. Therefore, regardless of the limitations presented, the machine learning algorithms were able to estimate with some level of accuracy the type of soil when both the electrical resistivity of the soil and its water content was given or when only the electrical resistivity of the soil was provided.

Chapter 6.

Conclusions and Future Work

6.1. Summary

In this dissertation, the electrical resistivity and soil moisture were used to predict two different types of soil using k nearest neighbors (KNN) and support vector machine algorithm (SVM). Besides predicting different types of soil, this research used the measured electrical resistivity and corrosion standards such as the American Water Works Association to make recommendations for materials that may withstand electrochemical corrosion within a hybrid environment (e.g. a metal buried underground in soil and moisture). The electrical resistivity was measured using Wenner's four electrodes technique. To create the KNN literature model, a dataset of 162 sample points was obtained from different literature wherein 151, and 11 points were used as training and testing points, respectively. To predict the laboratory soil, 26 sample points were obtained (corresponding to 130 measurements) wherein 6 points were added to the literature training dataset and 20 used as testing points.

The results showed that the SVM algorithm is unfit to predict the laboratory sample soils (clayey silt and clay, or soil 1 and 2, respectively) while using the training dataset provided to create the machine learning models. The SVM algorithm was unable to separate the training sample points into distinct classes, thus it could not classify different types of soils with high accuracy. In terms of error rate and accuracy, the SVM algorithm had an error rate of 44.1% and an accuracy of only 55.9 %. However, the k nearest neighbor algorithm proved to be capable of predicting the type of soil of an unknown soil when the electrical resistivity and moisture of the given soil are known. The model was tested using both literature and laboratory testing data. The model proved to fit the literature testing data better than the laboratory testing dataset. The reason is that the model became a litter bit biased to the literature dataset since it was used to create the model. The confusion matrix statistical rate obtained from both literature and experimental results are presented in Table 4.12. This table shows that the literature

testing data had accuracy, error rate, sensitivity, specificity, and precision of 82%, 18%, 75%, 86%, and 75% respectively. On the other hand, the experimental testing data points had accuracy, error rate, sensitivity, specificity, and precision of 70%, 30%, 64%, 83.3%, and 64% respectively. The k-NN model was capable of predicting the laboratory soils 1 and 2 (clayey silt and clay respectively). Additionally, the result of this dissertation showed that the model was not able to tell a mixture of soil from plain soil (e.g. clayey silt from clay).

In summary, the hypothesis of this dissertation states that given the electrical resistivity and moisture of an unknown soil, a machine learning algorithm can be leveraged to predict the type of soil. The implemented algorithm was capable of predicting the two types of soil used in the laboratory with some degree of accuracy, which can be built on for future implementations.

6.2. Recommendation for future work

There are several aspects of this dissertation that could be improved because some of the results presented were not conclusive. The limitation section of this research showed several aspects that could be improved in the future. Some of the aspects that could be improved or added in the future are the algorithm used to create the model, a physical version of the model that could be implemented, and field testing.

6.2.1. Model improvement

As previously mentioned, the data used to create the model were obtained from different scholars' papers. Although these scholar's papers are credible sources, there might be some calculations or other human mistakes that were not reported. These possible measurement errors committed by scholars affected negatively the model used in this research. However, to improve the accuracy of the model in the future, the data used to create the model will not be extracted from external sources, it will be obtained from a laboratory experiment conducted at SFU. This will increase the accuracy and improve the reliability of the model because all the errors and assumptions used during the data collection will be accounted for.

The model was created using about 162 data points wherein 80% of the data was used to train the model and about 20% to test the model. To improve the model accuracy in the future, more than 162 data points will be used. Since the data points used to create the model will be extracted from the laboratory at SFU, it would not be a big challenge to gather more than 162 points. In the future, a third soil should be used to further verify the hypothesis of this dissertation.

Identifying how much overfitting exists in a model is very difficult [38]. However, it is possible to analyze the result of our model to help us decide either the model is overfitting or not. The results of this research showed that the literature model over-performed the laboratory model with an accuracy of 82%, and 70% respectively. Statistically, it might be an indication of overfitting as previously mentioned. Overfitting might have happened because the literature model became too complex as the variance of the model increased at the training stage. Variance is the sensitivity of an algorithm to specific sets of the training dataset that occurs when the algorithm has limited flexibility to learn the true signal from the dataset [39]. As the literature variance increased, the model's ability to classify the training data increased as well, resulting in an accuracy of 82%. The increase of the variance, decreased the ability of the model to generalize when new datasets are input to the model. Therefore, the accuracy of the model decreased as we introduced unseen data (laboratory data) to the model, resulting in an accuracy decrease of 12%. To solve this problem, the concept of regularization will be used in the future. This technique is widely used to reduce the variance and error of the model, thus avoiding overfitting [40].

Finally, the model will be improved by the addition of electrodes depth of penetration and temperature as two keys parameters to determine the type of soil. By using four features to train the model instead of two the accuracy of the model will be greatly improved. Also, a new algorithm and software might be used instead of the MATLAB and just the k-NN algorithm and SVM.

6.2.2. Device implementation

As mentioned in the objective section of this dissertation, the model created in this research could be embedded within a multi-use hardware device that also includes ACPS, or implemented as a separate device. If the model discussed in this report is implemented as an additional feature of the ACPS, it would improve greatly the capability of the ACPS. The ACPS could utilize a machine-learning algorithm to predict the type of soil. Additionally, instead of just protecting the metal target, it would also help engineers to select optimal materials that may withstand electrochemical corrosion in a hybrid environment (soil and moisture). The machine learning feature would make the device more robust and unlike any other device available in the market, it would be able to conduct a complete investigation of the soil before placing a metallic structure underground. However, if the model discussed herein is implemented as a single device, it would be used to measure soil resistivity, soil moisture, and estimate the type of soil under investigation.

Since the device is intended to be portable, the device will have a rechargeable battery, embedded multimeter (hardware), and a user interface. The battery will supply the DC power to the Wenner's four electrodes. The user will have the ability to adjust the input current and read the voltage drop from the user interface. Additionally, the user should be able to read the electrical resistivity from the user interface. In terms of data storage, all the training data should be stored in the device using a non-volatile memory card. The training dataset will be obtained from several in-field measurement testing. The memory card (database) will be periodically updated with new data to improve the classification algorithm overtime.

6.2.3. Field testing

After the device implementation, the next phase of this research would be the field test to verify the performance of the device. The device would be tested in different sites to verify its accuracy to measure soil resistivity, estimate soil type, and moisture. By measuring soil resistivity accurately and estimate the type of soil the device's ability to leverage parameters such as electrical resistivity and soil moisture to help engineers to

determine the aggressiveness of the soil in terms of corrosion would be determined. Finally, by testing both known and unknown types of soil the device's ability to predict the type of soil in the field would be verified.

References

[1] D. N. KHOI, "DK NGOC KHOI Trading," DK NGOC KHOI TRADING & SERVICES CO., LTD, [Online]. Available: http://dkngockhoi.com/n/so-sanh-su-khac-nhau-giua-an-mon-dien-hoa-va-an-mon-hoa-hoc.html#:~:text=Chemical%20corrosion%20is%20the%20redox,the%20anode%20to%20the%20cathode.. [Accessed 20 12 2020].

[2] K. T. M. Y. N. T. K. Y. Itsuo Yamaura, "An Estimation Method of Ground Resistance of Trees Growing," IEEE Inshumentation and Measurement: Technology Conference, pp. 949-952, 2002.

[3] J. McNEIL, "Electrical conductivity of soils and rocks," GEONICS LIMITED, pp. 5-21, 1980.

[4] "Standard Test Method for Field Measurement of Soil Resistivity Using the Wenner Four-Electrode Method1," Designation: G57 − 06 (Reapproved 2012, pp. 1-6, 2012.

[5] T. K. K. Norsangsri, "Estimation of soil resistivity based on multispectral image classification," Power engineering letters , vol. 2, pp. 1280-1281, 2011.

[6] IAN, "greymatters," 18 09 2013. [Online]. Available: https://greymattersglobal.com/soil-resistivity-testing-methods-schlumberger-array/. [Accessed 07 05 2020].

[7] B. Philip, "Resistivity Testing for Earthing Safety," IEEE, pp. 1-16, 2016.

[8] T. H. Vivek Sai V, "Computational Methods for Simulating Soil Parameters using Electrical Resistivity Technique," IEEE, pp. 1-7, 2017.

[9] Q. B. Li Liangfu, "Research on Influence of Soil Water Content on Soil Resistivity," International Conference on Ligtning Protection , pp. 1-7, 2012.

[10] R. S. F. A. D. C. W. a. A. S. Y. Mohd Hazreek Zainal Abidin, "Soil Moisture Content and Density Prediction Using Laboratory Resistivity Experiment," IACSIT International Journal of Engineering and Technology, vol. 5, no. 6, pp. 1-6, 2013.

[11] A. K. H. S. Sajjad Ahmad *, "Estimating soil moisture using remote sensing data: A machine learning approach," Science Direct , pp. 69-80, 2010.

[12] M. L. M. L. a. C. W. Blandine Lemercier, "Extrapolation at regional scale of local soil knowledge using boosted classification," Science Direct , pp. 75-84, 2012.

[13] J. A. M.-C. M. H. S. J. M. I. E. N. T. a. A. G. Mohsen BAGHERI BODAGHABADI, "Digital Soil Mapping Using Artificial Neural Networks and Terrain-Related Attributes," Elsevier, pp. 580-591, 2015.

[14] A. Instrument, "Understanding Soil Resistivity Testing," AEMC Instrument , Massachusetts , USA, 2019.

[15] A. W. W. Association, "Polyethylene Encarsement for Ductile-Iron Pipe Systems," pp. 1-19, 2010.

[16] J. R. Davis, Corrosion: Understanding the Basics, Ohio: The Materials Information Society, 2000.

[17] N. Perez, Electrochemistry And Corrosion Science, Boston: Kluwer Academic Publishers , 2004.

[18] W. Solken, "EXPLORE the WORLD of PIPING," [Online]. Available: http://www.wermac.org/materials/corrosion.html. [Accessed 16 06 2020].

[19] I. G. A. S. O. O S I Fayomi, "Economic Impact of Corrosion in Oil Sectors and Prevention: An Overview," Jornal of Physics: Coference Series , pp. 1-8, 2019.

[20] L. a. S. Tecnology, "Earthing Fundamentals," Lightning and Surge Tecnology , pp. 2-40.

[21] F. Jones, "UBC Earth and Ocean Science," 28 06 2007. [Online]. Available: https://www.eoas.ubc.ca/ubcgif/iag/methods/meth_1/principles.htm. [Accessed 21 12 2020].

[22] R. S. F. A. D. C. W. a. A. S. Y. M. H. Z. Abidin, "Soil Moisture content and density prediction using laboratory resistivity experiment," ResearchGate , vol. 6, no. , pp. 731-735, 2013.

[23] F. Jones, "UBC Earth and Ocean Science," 28 06 2007. [Online]. Available: https://www.eoas.ubc.ca/ubcgif/iag/methods/meth_1/measurements.htm. [Accessed 18 12 2020].

[24] R. L. Enlightening, "THEORY of FOUR-ELECTRODE RESISTIVITY/CONDUCTIVITY METHOD," p. 4.

[25] I. The MathWorks, "The MathWorks," The MathWorks, Inc. , [Online]. Available: https://www.mathworks.com/discovery/machine-learning.html. [Accessed 29 06 2020].

[25] T. MathWorks, "MathWorks," The MathWorks, Inc. , [Online]. Available: https://www.mathworks.com/help/stats/nearest-neighbors-

1.html#:~:text=A%20Nearest%20neighbor%20search%20locates,%2C%20and%20Mahalanobis%2C%20among%20others.. [Accessed 23 06 2020].

[27] The MathWork, "MathWork," The MathWork,Inc., [Online]. Available: https://www.mathworks.com/help/stats/kdtreesearcher.html. [Accessed 23 06 2020].

[28] MathWorks, "MathWorks," The MathWorks, Inc, [Online]. Available: https://www.mathworks.com/help/stats/kdtreesearcher.html. [Accessed 07 09 2020].

[29] MathWorks, "MathWorks," The MathWorks, Inc, [Online]. Available: https://www.mathworks.com/help/stats/exhaustivesearcher.html. [Accessed 07 09 2020].

[30] T. MathWorks, "MathWorks," The MathWorks, Inc, [Online]. Available: https://www.mathworks.com/help/stats/exhaustivesearcher.knnsearch.html. [Accessed 07 09 2020].

[31] O. O. M. O. B. B. S. A. Ganiyu, "Investigation of soil moisture content over a cultivated farmland in Abeokuta Nigeria using electrical resistivity methods and soil analysis," Journal of King Saud University-Science , pp. 1-11, 2020.

[32] P. A. P. S. L. V. M. F. a. M. T. Brocca G. Calamita, "Electrical Resistivity and TDR methods for soil moisture estimation in central Italy testing-sites," Journal of Hydrology, pp. 1-12, 2012.

[33] S. R. A. F. D. C. W. M. I. M. A. T. B. M. F. A. F. K. Y. R. M. F. M. D. A. T. S. A. a. Z. M. H. Z A M Hazreek*, "Determination of Soil Moisture Content using Laboratory Experimental and Field Electrical Resistivity Values," Journal of Physics: Conference series , pp. 1-12, 2018.

[34] K. Markham, "Data School," 25 03 2014. [Online]. Available: https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/. [Accessed 05 11 2020].

[35] MathWorks, "The MathWorks," The MathWorks, Inc, [Online]. Available: https://www.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html. [Accessed 09 09 2020].

[36] MathWorks, "The MathWorks," The MathWorks, Inc, [Online]. Available: https://www.mathworks.com/help/stats/classificationecoc.html. [Accessed 07 09 2020].

[37] S. S. N. M. M. I. A. M. A. Muhammad Wasim, "Factors influencing corrosion of metal pipes in soils," Springer Link, p. 861–879, 2018.

[38] C. F. Institute, "CFI," [Online]. Available: https://corporatefinanceinstitute.com/resources/knowledge/other/overfitting/#:~:text=Overfitting%20can%20be%20identified%20by,model%20is%20affected%20by%20overfitting.. [Accessed 20 12 2020].

[39] A. Guanga, "Medium," 11 10 2018. [Online]. Available: https://becominghuman.ai/machine-learning-bias-vs-variance-641f924e6c57. [Accessed 20 12 2020].

[40] S. Biswas, "Medium," 8 5 2019. [Online]. Available: https://medium.com/towards-artificial-intelligence/how-regularization-can-help-in-overfitting-the-data-ad9ff80f9ccc. [Accessed 20 12 2020].

Appendix.

Supplemental Datasets

Table A1. k-NN Training and Testing Data

resistivity	Soil Moisture	Level of Corrosion (Powertech)	Description
1212.25	17.81	0	Munnar, India at a single layer @depth of 0.3m
448.4	0.266	0	Munnar, India at a single layer @depth of 0.3m
358.72	8.68	0	Munnar, India at a single layer @depth of 0.3m
306.8	12.09	0	Munnar, India at a single layer @depth of 0.3m
205.61	24.74	0	Munnar, India at a single layer @depth of 0.3m
195.88	26.84	0	Munnar, India at a single layer @depth of 0.3m
186.2	29.14	0	Munnar, India at a single layer @depth of 0.3m
147.33	42.08	0	Munnar, India at a single layer @depth of 0.3m
131.84	49.82	0	Munnar, India at a single layer @depth of 0.3m
44.96	31.7	0	CBE, site: sandy loam-Vallaccia, Italy @depth of 0.2m
103.04	30.8	0	COL, site: sandy loam-Vallaccia, Italy @depth of 0.2m
32.69	32.3	0	CON, site: sandy loam-Vallaccia, Italy @depth of 0.2m
62.02	31.9	0	CRI, site: silt loam-Vallaccia, Italy@depth of 0.2m
12.47	42.9	10	LEC, site: sandy loam-Vallaccia, Italy@depth of 0.2m
39.3	24.7	0	MOL, site: sandy-Vallaccia, Italy@depth of 0.2m
15.67	38.2	8	PRE, site: sandy loam-Vallaccia, Italy @depth of 0.2m
38.12	32	0	VRO, site: sandy loam-Vallaccia, Italy @depth of 0.2m
5600	6.36	0	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
1100	10.07	0	Clayey Silt-Universiti Tun Hussein Onn,Malaysia@depth of 0.24m
800	12.03	0	Clayey Silt-Universiti Tun Hussein Onn,Malaysia@depth of 0.24m
380	13.47	0	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
160	17.62	0	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
100	19.91	0	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
58	24.25	0	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
40	26.45	0	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
31	28.71	0	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
21	30.96	5	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m

resistivity	Soil Moisture	Level of Corrosion (Powertech)	Description
19	36.82	5	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
18	35.14	8	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
16	37.05	8	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
16	37.86	8	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
17	43.11	8	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
15	45.7	8	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
15	52.78	8	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
15	56.09	8	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
14	60.97	10	Clayey Silt-Universiti Tun Hussein Onn,Malaysia @depth of 0.24m
78	45	0	VES 1,site: sandy loam, Nigeria @depth of 0.5m
275	48	0	VES 1,site: sandy loam, Nigeria @depth of 1.6m
26	50	1	VES 1,site: sandy loam, Nigeria @depth of 2.7m
349	49	0	VES 2,site: sandy loam, Nigeria @depth of 1.1m
98	52	0	VES 2,site: sandy loam, Nigeria @depth of 6.8m
475	66	0	VES 2,site: sandy loam, Nigeria @depth of 6.8m
1094	51.5	0	VES 3,site: sandy loam, Nigeria @depth of 0.8m
893	52	0	VES 3,site: sandy loam, Nigeria @depth of 12.3m
393	50.2	0	VES 4,site: sandy loam, Nigeria @depth of 0.9 m
110	50.5	0	VES 4,site: sandy loam, Nigeria @depth of 4.8 m
766	71.5	0	VES 4,site: sandy loam, Nigeria @depth of 4.8 m
122	51	0	VES 5,site: sandy loam, Nigeria @depth of 1.3m
26	54	1	VES 5,site: sandy loam, Nigeria @depth of 15.3 m
282	55	0	VES 5,site: sandy loam, Nigeria @depth of 15.3 m
112	51	0	VES 6,site: sandy loam, Nigeria @depth of 1.9m
18	54	8	VES6,site: clay loam, Nigeria @depth of 10.5 m
181	55	0	VES6,site: clay loam, Nigeria @depth of 10.5 m
34.37	13.5	0	Meteorological Bureau of Hechuan District, China @depth of 0.05m
34.37	43.39	0	Meteorological Bureau of Hechuan District, China @depth of 0.10m
34.37	22.1	0	Meteorological Bureau of Hechuan District, China @depth of 0.20m
34.37	15.9	0	Meteorological Bureau of Hechuan District, China @depth of 0.30m
44.27	21.3	0	Meteorological Bureau of Hechuan District, China @depth of 0.50m

resistivity	Soil Moisture	Level of Corrosion (Powertech)	Description
69.66	18.6	0	Meteorological Bureau of Hechuan District, China @depth of 1m
85	10.1	0	Meteorological Bureau of Hechuan District, China @depth of 1.8m
355	60	0	Lower lias: Dark fibrous loam @ the surface ; Rugby Radio
750	33	0	Lower lias: Loam and clay @depth of 0.31m ; Rugby Radio
860	26	0	Lower lias: Clay and sand @depth of 0.61m ; Rugby Radio
945	25	0	Lower lias: Blue clay @depth of 0.91m ; Rugby Radio
695	23	0	Lower lias: Blue Clay @depth of 3.10m ; Rugby Radio
66.5	22	0	Lower lias: Loam @the surface ; Rugby Radio
94	13	0	Lower lias: Loam and clay @depth of 0.31m ; Rugby Radio
775	27	0	Lower lias: Blue clay @depth of 0.91 m ; Rugby Radio
610	21	0	Lower lias: clay and sand @depth of 1.52m ; Rugby Radio
835	25	0	Lower lias: Blue clay @depth of 3.10 m ; Rugby Radio
97	21	0	Chalk: Fibrous loam @ the surface; Tatsfield, Kent
61	21	0	Chalk: chalcky loam @ the depth of 0.31m; Tatsfield, Kent
55	24	0	Chalk: chalk @ the depth of 0.61m; Tatsfield, Kent
75	27	0	Chalk: chalk @ the depth of 0.91 m; Tatsfield, Kent
128.5	26	0	Chalk: chalk @ the depth of 1.52 m; Tatsfield, Kent
150	27	0	Chalk: chalk @ the depth of 3.1 m; Tatsfield, Kent
340	37	0	Upper green sand: Fibrous loam @ the surface; Tatsfield, Kent
255	11	0	Upper green sand: Brown, sand clay @ the depth of 0.31 m; Tatsfield, Kent
210	15	0	Upper green sand: Brown sand @ the depth of 0.61 m; Tatsfield, Kent
131.5	13	0	Upper green sand: Light brown sand @ the depth of 0.91 m; Tatsfield, Kent
260	20	0	Upper green sand: Light brown sand @ the depth of 1.52 m; Tatsfield, Kent
178.5	15	0	Upper green sand: Yellow sand @ the depth of 3.10 m; Tatsfield, Kent
103	19	0	London Clay: Fibrous loam @ the surface; Brookmans Park, Herts
97	18	0	London Clay: Stony loam @ the depth 0.31 m; Brookmans Park, Herts
64	22	0	London Clay: Light sand Clay @ the depth 0.61 m; Brookmans Park, Herts
127.5	22	0	London Clay: Sand clay @ the depth 0.91 m; Brookmans Park, Herts
177.5	21	0	London Clay: Sand clay @ the depth 1.52 m; Brookmans Park, Herts
172.5	10	0	London Clay: Clay and Shingle @ the depth 3.10 m; Brookmans Park, Herts
100	28	0	Upper lias: Fibrous loam @ the suface; Daventry Northants

resistivity	Soil Moisture	Level of Corrosion (Powertech)	Description
38	16	0	Upper lias: Sand loam @ the depth of 0.31m; Daventry Northants
32	14	0	Upper lias: Brown sand @ the depth of 0.61m; Daventry Northants
11.5	5	10	Upper lias: Brown sand @ the depth of 0.91m; Daventry Northants
9	8.5	10	Upper lias: Sand and sandstone @ the depth of 1.52m; Daventry Northants
33	24	0	Upper lias: Sand and sandstone @ the depth of 3.10m; Daventry Northants
178	23	0	Red Marls: Reddish-brown loam @ the surface ; Washford Cross, Somerset
178	20	0	Red Marls: Reddish-brown clay @ the depth of 0.31 m ; Washford Cross, Somerset
299.5	18	0	Red Marls: Reddish-brown clay @ the depth of 0.61 m ; Washford Cross, Somerset
530	21	0	Red Marls: Reddish-brown clay @ the depth of 0.91 m ; Washford Cross, Somerset
325	19	0	Red Marls: Reddish-brown clay @ the depth of 1.52 m ; Washford Cross, Somerset
490	15	0	Red Marls: Reddish-brown clay @ the depth of 3.10 m ; Washford Cross, Somerset
155.5	21	0	Devonian: Black fibrous loam @ the surface; Brendon Hills, Somerset
3	9	10	Devonian: Loam and slate @ the depth of 0.31m; Brendon Hills, Somerset
2	9	10	Devonian: Loam and slate @ the depth of 0.61m; Brendon Hills, Somerset
1	8	10	Devonian: Loam and slate @ the depth of 0.91 m; Brendon Hills, Somerset
0	5.5	10	Devonian: Loam and slate @ the depth of 1.52m; Brendon Hills, Somerset
0	0	10	Devonian: Slate @ the depth of 3.10m; Brendon Hills, Somerset
13	18	10	Granite: Gritty loam @ the surface of 0.31m; Merrivale, Dartmoor,Devon
16	13	8	Granite: Gritty loam @ the surface of 0.61m; Merrivale, Dartmoor,Devon
0.5	0	10	Granite: Granite @ the surface of 1.22m; Merrivale, Dartmoor,Devon
0.5	0	10	Granite: Granite @ the surface of 1.83m; Merrivale, Dartmoor,Devon
0	0	10	Granite: Granite @ the surface of 2.73m; Merrivale, Dartmoor,Devon
64	47	0	Devonian: Loam @ the surface; Dousland, Dartmoor, Devon
27	41	1	Devonian: Dark brown laom @ the depth of 0.31 m ; Dousland, Dartmoor, Devon
0	0	10	Devonian: Slate @ the depth of 0.1 m; Dousland, Dartmoor, Devon
0	0	10	Devonian: Granite @ the depth of 0.2 m; Dousland, Dartmoor, Devon

resistivity	Soil Moisture	Level of Corrosion (Powertech)	Description
155.5	13	0	Millstone grit: Fibrous loam @ the surface; Moorside, Edge, Yorks
144.5	60	0	Millstone grit: Dark grey clay @ the depth of 0.31m; Moorside, Edge, Yorks
92	35	0	Millstone grit: Dark grey clay @ the depth of 0.61m; Moorside, Edge, Yorks
144.5	39	0	Millstone grit: Dark grey clay @ the depth of 0.91m; Moorside, Edge, Yorks
102.5	19	0	Millstone grit: Dark grey clay @ the depth of 1.52m; Moorside, Edge, Yorks
35.5	15	0	Millstone grit: Yelow and grey clay @ the depth of 3.10m; Moorside, Edge, Yorks
66.5	38	0	Boulder clay: Fibrous loam @ the surface; Westerglen, Falkirk
111	30	0	Boulder clay: Fibrous loam @ the depth of 0.31m; Westerglen, Falkirk
122	19	0	Boulder clay: Clay and loam @ the depth of o.61m; Westerglen, Falkirk
72.5	18	0	Boulder clay: Dark grit and clay @ the depth of 0.91m; Westerglen, Falkirk
225	18	0	Boulder clay: Dark grit and clay @ the depth of 1.52m; Westerglen, Falkirk
245	15	0	Boulder clay: Dark grit and clay @ the depth of 3.10m; Westerglen, Falkirk
116.5	26	0	London clay: Fibrous loam @ the surface; Teddington, Middlesex
75	20	0	London clay: Sandy loam @ the depth of 0.31m; Teddington, Middlesex
69.5	13	0	London clay: Sandy loam @ the depth of 0.61m; Teddington, Middlesex
61.5	6.5	0	London clay: Fine gravel @ the depth of 0.91m; Teddington, Middlesex
19.5	2.9	5	London clay: Coarse gravel @ the depth of 1.52m; Teddington, Middlesex
16	2.6	8	London clay: Fine sand @ the depth of 2.13m; Teddington, Middlesex
144.5	20	0	London clay: Sand and Shingle @ the depth of 3.10 m; Teddington, Middlesex
183.5	15	0	Red Marls: Red clay and loam @ depth of 0.31m; Wychbold, Droitwith
1.885	30	10	unknown experimental soil_1 @ depth of 0.01 m
11.2762	30	10	unknown experimental soil_1 @ depth of 0.02 m
3.3091	70	10	unknown experimental soil_2 @ depth of 0.01 m
23.2897	35	2	unknown experimental soil_2 @ depth of 0.02 m
5.3617	70	10	unknown experimental soil_2 @ depth of 0.02 m
51.6059	35	0	unknown experimental soil_2 @ depth of 0.04 m

Table A2. SVM Training Data

Resistivity	Moisture	Level of Corrosion	Description
62.02	31.9	0	Silt loam
12.47	42.9	10	Sandy loam
39.3	24.7	0	Sandy loam
15.67	38.2	8	Sandy loam
38.12	32	0	Sandy loam
380	13.47	0	Clayey Silt
160	17.62	0	Clayey Silt
31	28.71	0	Clayey Silt
21	30.96	5	Clayey Silt
19	36.82	5	Clayey Silt
18	35.14	8	Clayey Silt
16	37.05	8	Clayey Silt
16	37.86	8	Clayey Silt
17	43.11	8	Clayey Silt
15	45.7	8	Clayey Silt
15	52.78	8	Clayey Silt
15	56.09	8	Clayey Silt
14	60.97	10	Clayey Silt
1094	51.5	0	Sandy loam
893	52	0	Sandy loam
393	50.2	0	Sandy loam
110	50.5	0	Sandy loam
766	71.5	0	Sandy loam
122	51	0	Sandy loam
26	54	1	Sandy loam
282	55	0	Sandy loam
112	51	0	Sandy loam
181	55	0	Clay loam
355	60	0	Dark fibrous loam
750	33	0	Loam and clay
860	26	0	Clay and sand
94	13	0	Loam and clay
610	21	0	Clay and sand
835	25	0	Blue clay
97	21	0	Fibrous loam
61	21	0	Chalcky loam
55	24	0	Chalk
128.5	26	0	Chalk
150	27	0	Chalk

Resistivity	Moisture	Level of Corrosion	Description
260	20	0	Light brown sand
178.5	15	0	Yellow sand
103	19	0	Fibrous loam
97	18	0	Stony loam
64	22	0	Light sand Clay
127.5	22	0	Sand clay
177.5	21	0	Sand clay
172.5	10	0	Clay and Shingle
100	28	0	Fibrous loam
38	16	0	Sand loam
32	14	0	Brown sand
11.5	5	10	Brown sand
33	24	0	Sand and sandstone
530	21	0	Reddish-brown clay
325	19	0	Reddish-brown clay
490	15	0	Reddish-brown clay
155.5	21	0	Black fibrous loam
0	0	10	Slate
13	18	10	Gritty loam
16	13	8	Gritty loam
0.5	0	10	Granite
0.5	0	10	Granite
0	0	10	Granite
64	47	0	Loam
27	41	1	Dark brown laom
0	0	10	Slate
0	0	10	Granite
155.5	13	0	Fibrous loam
92	35	0	Dark grey clay
144.5	39	0	Dark grey clay
102.5	19	0	Dark grey clay
35.5	15	0	Yelow and grey clay
66.5	38	0	Fibrous loam
111	30	0	Fibrous loam
122	19	0	Clay and loam
72.5	18	0	Dark grit and clay
225	18	0	Dark grit and clay
245	15	0	Dark grit and clay
116.5	26	0	Fibrous loam
75	20	0	Sandy loam

Resistivity	Moisture	Level of Corrosion	Description
69.5	13	0	Sandy loam
61.5	6.5	0	Fine gravel
19.5	2.9	5	Coarse gravel
16	2.6	8	Fine sand
144.5	20	0	Sand and Shingle
183.5	15	0	Red clay and loam
51.6059	35	0	unknown experimental soil_2
9.93	10	10	unknown experimental soil_2
11.2762	30	10	unknown experimental soil_1
32.69	32.3	0	Sandy loam
5600	6.36	0	Clayey Silt
1100	10.07	0	Clayey Silt
800	12.03	0	Clayey Silt
78	45	0	Sandy loam
275	48	0	Sandy loam
3	9	10	Loam and slate
2	9	10	Loam and slate
26	50	1	Sandy loam
340	37	0	Fibrous loam
255	11	0	sand clay
210	15	0	Brown sand
1	8	10	Loam and slate
0	5.5	10	Loam and slate
1.885	30	10	unknown experimental soil_1
3.3091	70	10	unknown experimental soil_2
23.2897	35	2	unknown experimental soil_2
5.3617	70	10	unknown experimental soil_2
349	49	0	Sandy loam
100	19.91	0	Clayey Silt
58	24.25	0	Clayey Silt
40	26.45	0	Clayey Silt
98	52	0	Sandy loam
475	66	0	Sandy loam
75	27	0	Chalk
178	20	0	Reddish-brown clay
144.5	60	0	Dark grey clay
18	54	8	Clay loam
945	25	0	Blue clay
131.5	13	0	Light brown sand
9	8.5	10	Sand and sandstone

Resistivity	Moisture	Level of Corrosion	Description
178	23	0	Reddish-brown loam
299.5	18	0	Reddish-brown clay
12.4	65	10	unknown experimental soil_2
21.82	70	2	unknown experimental soil_2
38.03	42	0	unknown experimental soil_2
6.64	35	10	unknown experimental soil_2
18.56	90	5	unknown experimental soil_1
5.45	65	10	unknown experimental soil_2
53.99	10	0	unknown experimental soil_2

Table A3. Soil 1-at 1 cm depth

Voltage drop trials (V)					Current (A)					R (Ω)	a (m)	Temper-ature	Soil Moisture	P (Ω · m)
1.3	0.3	1	0.6	0.7	0.03	0.02	0.03	0.03	0.03	27	0.02	28	35	3.40
1.2	1.5	1.6	1.4	1.3	0.04	0.04	0.04	0.04	0.04	35	0.02	27	90	4.40

Table A4. Soil 1-at 2 cm depth

Voltage drop trials (V)					Current (A)					R (Ω)	a (m)	Temper-ature	Soil oisture	P (Ω · m)
1.8	1.2	1.7	1.6	1.8	0.03	0.03	0.03	0.03	0.03	54	0.04	28	35	13.57
3.1	3.1	3.1	3.1	3.1	0.05	0.06	0.06	0.04	0.05	61	0.04	27	90	15.32

Table A5. Soil 2-at 1 cm depth

Voltage drop trials (V)					Current (A)					R (Ω)	a (m)	Temper-ature	Soil Moisture	P (Ω · m)
1.8	1.8	2.2	1.8	1.2	0.03	0.03	0.03	0.03	0.03	58.7	0.02	28	40	7.37
1.4	1.5	1.4	1.5	1.4	0.03	0.03	0.03	0.03	0.03	48	0.02	28	80	6.03

Table A6. Soil 2-at 2 cm depth

Voltage drop trials (V)					Current (A)					R (Ω)	a (m)	Temper-ature	Soil Moisture	P (Ω · m)
1.6	1.7	1.9	1.9	2.1	0.03	0.03	0.03	0.03	0.03	61.33	0.04	28	40	15.41
1.9	1.2	1.8	1.4	1.8	0.03	0.03	0.03	0.03	0.03	54	0.04	27	80	13.57

Table A7. Soil 2-at 4 cm depth

Voltage drop trials (V)					Current (A)					R (Ω)	a (m)	Temper-ature	Soil Moisture	P (Ω · m)
2.4	2.2	2.2	2.3	2.6	0.03	0.03	0.03	0.03	0.03	78	0.08	28	40	39.20
3.2	2.7	2.8	2.9	3.3	0.03	0.04	0.03	0.03	0.03	94.83	0.08	27	80	47.67