TO STUDY THE QUALITY OF SOIL AND ITS INDICATORS USING MACHINE LEARNING IN KANPUR AGRICULTURAL LANDS.

1Mr. Avinash Bajpai*, 2 Dr. Jitendra Nath Shrivastava, 3Dr. Swati Saxena, 4Dr. Pratibha Mishra

1Research Scholar, INVERTIS UNIVERSITY, Bareilly (UP), India
2 Professor, INVERTIS UNIVERSITY, Bareilly (UP), India
3 Assistant Professor, Maharana Pratap Group of Institutions, Kanpur, India
4 Assistant Professor, DAV PG COLLEGE, Kanpur, India

ABSTRACT

Precision agriculture can be defined as the art and science of using advanced technology to enhance crop production. Machine learning using Image processing is a major technology that drives the development of precision agriculture.

The proposed analysis increase the agriculture growth on the basis of decision through machine learning using image processing using low cost technology in day to day life to increase the productivity of the agriculture.

To study the quality of soil and its indicator using machine learning that can measured to monitor changes in the soil. Soil quality indicators are physical, chemical, and biological properties, processes, and characteristics that can be measured to monitor changes in the soil. Soil quality indicators are important to: focus conservation efforts on maintaining and improving the condition of the soil; evaluate soil management practices and techniques; relate soil quality to that of other resources; collect the necessary information to determine trends; determine trends in the health of the Nation’s soils; guide land manager decisions.

The main aim of this paper is to review current efforts to define soil quality, and to discuss factors and processes which influence soil quality, to identify, soil and crop management practices using machine learning that affect processes influencing soil quality, and to demonstrate a method for evaluating soil quality. A common focus among all proposed soil quality definitions is that the soil must reflect its ability to “function” in numerous ways at the present time and in the future. Soil and crop management using
machine learning practices that add or maintain soil carbon appear to be among the most important for restoring, maintaining, or improving soil quality. The soil quality assessment method that has been developed does not provide a definitive answer with regard to the measurements or specific functions which should be included in a soil quality index, but it uses specific measurements that describe soil functions and it is dynamic. Measurement of decomposition rates of plant residue in bags or measurements of weed seed numbers, or pathogen populations can serve as biological indicators of soil quality.

Physico-chemical and biological properties of soils from different land use systems viz. agriculture, horticulture etc in Kanpur Uttar Pradesh were analyzed in 2021. Samples were collected from 4 different sites of Kanpur district viz., Bithoor, Tikra, IIPR & C.S.A having distance between them of at least 4 kms. Some soils samples had higher water holding capacity & value of organic carbon. They have average value of pH and EC as compared to the cultivated soils. Potassium was found to be of low content in soil samples. Physical properties and parameters for all soils were average or medium whereas as variation in chemical properties were observed.

**Key words:** Image Processing, Machine learning, Soil quality, soil functions, soil indicators.

**INTRODUCTION**

Indicators of soil quality can be categorized into four general groups: visual, physical, chemical, and biological.

Visual indicators may be obtained from observation or photographic interpretation and image processing. Exposure of subsoil, change in soil color, ponding, runoff, plant response, weed species, blowing soil, and deposition are only a few examples of potential locally determined indicators. Image evidence can be a clear indication that soil quality is threatened or changing.

Physical indicators are related to the arrangement of solid particles and pores. Examples include top soil depth, bulk density, porosity, aggregate stability, texture, crusting, and compaction. Physical indicators primarily reflect limitations to root growth, seedling emergence, infiltration, or movement of water within the soil profile.

Chemical indicators include measurements of pH, salinityorganic matter, phosphorus concentrations, cation-exchange capacity, nutrient cycling, and concentrations of elements that may be potential contaminants (heavy metals, radioactive compounds, etc.) or those that are needed for plant growth and development. The soil’s chemical condition affects soil-plant relations, water quality, buffering capacities, availability of nutrients and water to plants and other organisms, mobility of contaminants, and some physical conditions, such as the tendency for crust to form.

Biological indicators include measurements of microand macro-organisms, their activity, or byproducts.

The concept of soil quality has been suggested by several authors (Lal, 1991; Granatstein and Bezdicek, 1992; Sanders, 1992; Karlen et al., 1992; Papendick and Parr, 1992; Parr et al., 1992; Acton and Padbury, 1993) as a tool for assessing long-term sustainability of agricultural practices at local, regional, national, and international levels. This suggestion was reinforced by a recent report from the National Academy of Sciences, National Research Council (1993) recommending that the United States adopt a national policy which seeks to conserve and enhance soil quality as a fundamental first step to environ-mental improvement. The main aim of this work is (1) to review current efforts to define soil quality; (2) to
discuss factors and processes which influence soil quality; (3) to identify soil and crop management practices that affect processes influencing soil quality; and (4) to demonstrate a potential method for evaluating soil quality.

Acton and Padbury (1993) proposed that the definition of soil quality should be based on two critical soil functions, each representing major expectations placed on soils by farmers and agricultural or other resource managers. These functions are (1) to ensure sustainable crop production or the capacity to produce crops; and (2) to ensure environmental sustainability or the capacity of soil to serve as an environmental buffer, to accept, hold and release water to plants, streams, and groundwater, and to function as a source or sink for gaseous materials and the capacity to exchange those materials with the above ground atmosphere. With this general background, several factors and processes which may influence soil quality will be examined.

**SITE SELECTION**

Soil samples, all laboratory chemical reagents, apparatus etc. were used for physical and chemical analysis of soil samples. Soil samples were collected from four regions (Atleast 4 km Apart from each other as suggested) namely Bithoor, Tikra, IIPR & C.S.A Kanpur. Chemicals were used from UG & PG LABS of CSA University. Soil samples were dried, powdered and sieved. Then these samples were taken for analysis of physical parameters viz. Bulk density, particle density, pore size, water retaining capacity, specific gravity etc. The samples were also analyzed for chemical parameters like pH, electrical conductivity, percentage organic carbon and potassium content.

Four sampling areas of 1 hectar each have been selected for testing of soil

Sample 1 was taken Bithoor with Latitude-26.6057° North & Longitude80.2698° East, Sample 2 was taken Tikra with Latitude 26.4978° North & Longitude80.2070° East,

Sample 3 was taken IIPR with Latitude26.4940° North & Longitude80.2720° East,

Sample 4 was taken C.S.A with Latitude26.4912° North & Longitude 80.3071° East.

All samples were collected from a depth of furrow slice layer. With respect to Climatic conditions of Kanpur all four samples were collected having dry weather conditions with a lit bit of temperature up to 28°C. We have worked on physical & chemical properties of the respective soil samples collected and before proceeding to experimental processes and data, we must see in general some basic properties as classified further. Physical properties of soil include Particle density, Bulk density, Percent pore space, Specific gravity, Soil color, Waterretaining capacity, etc. Whereas Chemical properties involve pH of soil, Electrical conductivity (EC), Percentage organic carbon, Potassium Estimation, etc.

**FACTORS SHOWING SOIL QUALITY**

Karlen et al. (1992) stated that inherent interactions among the five basic soil forming factors [parent material, climate (including water and temperature effects), macro- and micro-organisms, topography and time] identified by Jenny (1941) create a relatively stable soil quality that has distinct physical, chemical, and biological characteristics in response to prevailing natural or non-anthropogenic factors. However, humankind, the anthropogenic force described as a sixth soil forming factor in the basic model for describing a soil (SSSA, 1987), interacts with the non- anthropogenic factors and influences soil quality both negatively and positively. Soil and crop management practices imposed on land resources by humankind thus determine whether inherent soil quality will be lowered, sustained, or improved over
relatively short time intervals. The relative importance of anthropogenic or management factors compared to non-anthropogenic physical, chemical, or biological factors will generally be determined by the function or application for which a soil quality assessment is made.

Use of a minimum data set (MDS) for assessing the health or quality of world soils was proposed by Larson and Pierce (1991). They suggested that standardized methodologies and procedures be established to assess changes in soil quality. Soil attributes and measurements selected for their MDS (Table 1) were dictated by a need to be (1) sensitive to various soil and crop management practices; (2) detectable following relatively short. But variable periods of time; and(3) accessible to most people through direct measurement or pedotransfer functions (Bouma, 1989).

Doran and Parkin (1994) adapted the MDS recommendations and proposed several soil physical, chemical, and biological characteristics that should be included as basic indicators of soil quality (Table 2). They also provided a rationale for selecting these characteristics, and emphasized the importance of defining ecosystem mechanisms and control processes that respond to soil and crop management practices and ultimately determine soil quality.

Kay (1990) described soil structure in terms of form, stability and resiliency. Structural form describes the heterogeneous arrangement of solid and void space that exists at any given time and refers to the (1) arrangement of primary soil particles into hierarchical structures; (2) total porosity; (3) pore size distribution; and (4) continuity of the pore system. Soil stability is defined as the ability to retain solid and void space arrangement when exposed to different stresses such as compaction. Resiliency has not been specifically used in relation to soil structure, but Kay (1990) suggested that it provides a single term to describe processes such as tilth-mellowing, self-mulching, and age-hardening.
Table 1. Factors Recommended by Larson and Pierce (1991) for Inclusion in a Minimum Dataset for Assessment and Monitoring of Soil Quality.

<table>
<thead>
<tr>
<th>Texture or particle size</th>
<th>Pipette or hydrometer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil structure</td>
<td>Bulk density using intact cores or from water retention curves</td>
</tr>
<tr>
<td>Soil strength</td>
<td>Bulk density or penetrate on resistance</td>
</tr>
<tr>
<td>Maximum rooting depth or soil volume above root restrictive layers</td>
<td>On-site characterization for various crops or standard rooting estimates</td>
</tr>
<tr>
<td>Plant available water retention</td>
<td>Field measurements or estimation from water retention curves</td>
</tr>
<tr>
<td>Soil acidity or pH</td>
<td>pH meter with glass and reference electrodes</td>
</tr>
<tr>
<td>Electrolytic conductivity</td>
<td>Conductivity meter</td>
</tr>
<tr>
<td>Nutrient availability</td>
<td>Analytical soil test procedures (perhaps plant tissue analyses)</td>
</tr>
<tr>
<td>Total organic carbon (C)</td>
<td>Dry- or wet-combustion techniques</td>
</tr>
<tr>
<td>Lable organic C</td>
<td>CO₂-C release from hot KCl digests</td>
</tr>
</tbody>
</table>

Soil and crop management practices such as reduced tillage, increased input of carbon, and reduced pesticide applications may promote earthworm diversity and thus enhance the effects of earthworms on soil properties. Management practices that include polycultures, crop rotations, hedgerows, buffer strips, or reduced tillage may favor biodiversity and result in a number of benefits including an increased abundance of predators and beneficial parasites, and provide increased microhabitat diversity for microbial activity and processes (Hendrix et al., 1992).

**SOIL QUALITY EVALUATION**

Evaluating soil quality is difficult because it is much more site- and soil-specific than air or water quality. To meet this challenge, Larson and Pierce (1991) proposed five soil quality attributes, and suggested that the combined physical, chemical, and biological properties of a soil enable it to perform three functions. The soil functions are (1) to provide a medium for plant growth, (2) to regulate and partition water flow through the environment, and (3) to serve as an environmental filter. They also stated that soil quality describes how effectively soils:

1. Accept, hold, and release nutrients and other chemical constituents;
2. Accept, hold, and release water to plants, streams, and groundwater;
3. Promote and sustain root growth;
4. Maintain suitable soil biotic habitat; and
5. Respond to management and resist degradation.

Karlen and Stott (1994) proposed a framework for evaluating soil quality relative to water erosion that was based on soil processes and properties that were sensitive to soil and crop management practices. They identified four critical functions as (1) accommodating water entry into the soil, (2) facilitating water transport and absorption, (3) increasing resistance to soil erosion, and (4) supporting plant growth.

In subsequent studies, Karlen et al. (1994a) modified the framework to assess surface soil quality as affected by various crop residue and tillage treatments. Each biological, chemical, or physical measurement
that was used to compute the soil quality index was normalized to a value between 0 and 1 using standardized scoring functions.

Table 2. Soil physical, chemical, and biological characteristics proposed by Doruland Park in (1994) as basic indicators of soil quality.

<table>
<thead>
<tr>
<th>Soil characteristics</th>
<th>Relationship to soil condition or function</th>
<th>Rationale for selection as priority measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil texture</td>
<td>Retention and transport of water and chemical</td>
<td>Process modeling, erosion, and productivity estimates</td>
</tr>
<tr>
<td>Profile, topsoil and rooting depth</td>
<td>Productivity and erosion estimates</td>
<td>Normalization of landscape and geographic variables</td>
</tr>
<tr>
<td>Bulk density and water infiltration</td>
<td>Leaching, productivity, and erosivity estimates</td>
<td>Physical characteristics and for adjustment of measurements to volumetric basis</td>
</tr>
<tr>
<td>Water retention capacity</td>
<td>Water retention, transport, and erosivity</td>
<td>Water available for plant and microbial processes</td>
</tr>
<tr>
<td>Total organic C and N</td>
<td>Soil fertility, stability, and erosion status</td>
<td>Process modeling and normalization of site characteristics</td>
</tr>
<tr>
<td>pH</td>
<td>Biological and chemical activity thresholds</td>
<td>Process modeling</td>
</tr>
<tr>
<td>Electrical conductivity</td>
<td>Plant and microbial activity thresholds</td>
<td>Productivity and environmental quality indicators</td>
</tr>
<tr>
<td>Microbial biomass C and N</td>
<td>Microbial catalytic potential and capacity for C and N retention</td>
<td>Process modeling and surrogate indicator for microbial biomass</td>
</tr>
<tr>
<td>Potentially mineralizable N</td>
<td>Microbial and sometimes plant content and temperature activity</td>
<td>Process modeling and estimate of microbial biomass activity</td>
</tr>
</tbody>
</table>

(Wymore, 1993). The values chosen to normalize each soil quality measurement were derived from literature values for each parameter. Values selected for normalizing soil aggregation data were based on studies by Wilson and Browning (1945), while those for bulk density were as suggested by Singh et al. (1992) for their tilth index. Water-filled pore space normalization was based on information published by Doran et al. (1990) and Linn and Doran (1984a,b). For plant available water in silt loam soils, we utilized relationships suggested by Hudson (1993). Total carbon and total nitrogen scaling were based on experience with Rozetta and Palsgrove silt loam soils, while cation exchange capacity, microbial biomass, respiration, ergosterol concentrations, and earthworm populations were normalized based on literature reviewed by Eash (1993). After normalizing or scoring each measurement used for the proposed soil quality index, scores were multiplied by the appropriate weighting factor (Table 2). The products were then summed to give a weighted value. For factors such as nutrient relationships, weighted values for nutrient cycling (level 3) were computed and then used as the “score” for that factor at level 2. Similarly, all level 2 factors (pH, CEC, total N, total C, and nutrient cycling) were then multiplied by their respective weighting factor so that products could be summed to give weighted scores for each level 1 factor.
Weighted scores for each function were then summed to give an overall soil quality index as shown in another elements; supporting biological activity and diversity for plant and animal productivity; filtering, buffering, degrading, immobilizing, and detoxifying organic and inorganic materials; and providing mechanical support for living organisms and their structures.

Equation [1].

Soil Quality \( (Q) = \%\text{we} \times (\text{wt}) + \%\text{wta} \times (\text{wt}) + \%\text{rd} \times (\text{wt}) + \%\text{spg} \times (\text{wt}) \)

Where:
- \( \%\text{we} \) = Level I rating for accommodating water entry
- \( \%\text{wta} \) = Level I rating for water transport and absorption
- \( \%\text{rd} \) = Level I rating for resisting degradation
- \( \%\text{spg} \) = Level I rating for supporting plant growth
- wt = Weighting factor for each factor

Soil quality is estimated by observing or measuring several different properties or processes. No single property can be used as an index of soil quality. The selection of indicators should be based on: the land use; the relationship between an indicator and the soil function being assessed; the ease and reliability of the measurement; variation between sampling times and variation across the Karlen and Stott (1994) demonstrated how a soil quality index might be calculated using data from a study comparing alternate and conventional farming practices. The alternative farming practices, which included a 5-year corn, soybean, corn, oats, and meadow rotation; application of a mixture of animal manure and municipal sludge during the first 3-years of each rotation; and the use of ridge-tillage, resulted in a higher soil quality rating (0.73) than conventional practices (0.54), which consisted of a 2-year corn-soybean rotation without carbon input in excess of the crop residues. Using the framework shown in Table 2, Karlen et al. (1994a) computed soil quality index values showing that removal, maintenance, or doubling crop residues for 10 years with no-till production practices resulted in ratings of 0.45, 0.68, and 0.86, respectively. In another study (Karlen et al., 1994b), the same procedure indicated that the surface soil quality ratings after 10 years of plow, chisel, and no-till treatments were 0.47, 0.48, and 0.70, respectively. The relative ranking of the plow and no-till treatments in this study was confirmed by measuring soil loss with a sprinkling infiltrometer.

These initial studies have demonstrated the feasibility of developing a useful and perhaps valuable procedure for assessing surface soil quality. The procedure appears to be sensitive and can discern long-term crop residue management and tillage treatment effects. The proposed soil quality assessment method, although tested only for non-glaciated silt loam soils, does not provide a definitive answer with regard to measurements or specific functions which should be included in a soil quality index. However, it is based on actual soil measurements that describe specific soil functions and provides a framework for an even more dynamic soil quality index. Development of soil quality concepts is warranted and should enhance our efforts to achieve a more sustainable agriculture and environment.
ANALYSIS THROUGH MACHINE LEARNING

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

There are three types to train a machine:

- **Supervised Learning**: In this type of learning, machine is trained using the data where in we provide questions and answers for them to analyse and train. When the similar question is raised then the machine would output the stored answer.

  Example of one such type is Chabot, [1] where the questions and answers are given while training and if the similar queries are asked the given answers would be the output in the form of text or audio.

- **Unsupervised Learning**: In this type of learning, machines are left to themselves with the input data but the corresponding output is not known. This is like learning without guidance. Here these algorithms will come out with some interesting conclusions.

  Example of one such type is visual recognition and text recognition where the text from the given data is read and the conclusions or the related search takes place.

- **Reinforcement Learning**: We have no input or output to learn from this type of machine Learning technique. The question we raise is the input and the feedback/credit system (+,-) is the output to learn.

  Example of one such type is Tesla car, where the cars learn driving on their own (automated driving) and the rewards or credits or feedbacks are their training models.

**Requirement Specification**

The system was recognized with the following modules after thorough evaluation.

- **a. Dataset Module:**
  
  This module consists of the information regarding the datasets which are given as Input to the next modules. The datasets from excel sheet are imported and converted as a raw data for the input in the further procedure.

- **b. Training Module:**
  
  This module is the essential part of the whole system as the whole system is trained with machine learning algorithms and train the machine to perform the required task on its own without human intervention.

- **c. Testing Module:**
  
  This module is used to test the project and verify if the training has to be modified to get the appropriate precision in the output. Several graphs and matrix forms are used to make the user understand.
Figure 1

Machine Learning

Training set

Unsupervised

Feature extraction

Machine learning algorithm

Grouping of objects

Predictive model

New data

Annotated data

Figure 2

Phase 1: Define

Define Measure Analyze Improve Control

Define the problem, improvement activity, opportunity for improvement, the project goals, and customer requirements.

Use Design Thinking techniques if appropriate.

Key:
-.Tasks
- Outputs
PROPOSED SYSTEM IMPLEMENTATION

The proposed system uses different algorithms in machine learning for detecting physico-chemical analysis of soil in Kanpur agriculture land.

Algorithm

A) Decision Tree Algorithm: It is a supervised learning algorithm that is mostly used for classification problems. It works for either discrete or ongoing variables. A decision tree is plotted with its root at the top and branches at the bottom. The picture's courageous text reflects an internal condition / node on the basis of which the tree divides into branches / edges [42].

B) Random Forest Algorithm: For a supervised ensemble, Random Forest is a popular learning algorithm. Ensemble means that working together to form a strong predictor involves a lot of weak learners. In this scenario, the weak learners who are brought together to form the strong predictor—a random forest—are all randomly introduced choice trees. Like decision trees, trees forests also apply to issues with multiple outputs (if Y is a size range [ n samples, n outputs ]). Random Forest is a trade-mark term for a collection of decision trees. In Random Forest, we have a collection of decision trees called "Forest."[43] Each tree offers a scheme for the classification of a new item based on features and we say the tree "votes" for that class. The forest chooses the classification of the most votes (over all forest trees).

CONCLUSION

The types of indicators that are the most useful depend on the function of soil for which soil quality is being evaluated. These functions include: providing a physical, chemical, and biological setting for living organisms; regulating and partitioning water flow, storing and cycling nutrients and sampling area; the sensitivity of the measurement to changes in soil management; compatibility with routine sampling and monitoring; the skills required for use and interpretation.

Interpreting indicator measurements to separate soil quality trends from periodic or random changes is currently providing a major challenge for researchers and soil managers. Soils and their indicator values vary because of differences in parent material, climatic condition, topographic or landscape position, soil organisms, and type of vegetation. For example, cation exchange capacity may relate to organic matter, but it may also relate to the kind and amount of clay.

REFERENCES


