**PERFORMANCE OF SVM CLASSIFIER BASED SOIL CLASSIFICATION**

**ABSTRACT**

Quantitative techniques for prediction and classification in soil survey are developing rapidly. Classification of soil is the dissolution to soil sets to particular group having a like characteristics and similar manners. Almost all countries do product exporting, in which those countries exporting higher agricultural product are very much depend on the soil characteristics. Thus, soil characteristics identification and classification is very much important. Identification of the soil type helps to avoid agricultural product quantity loss. A classification for engineering purpose should be based mainly on mechanical properties. This paper explains support vector machine based classification of the soil types. Soil classification includes steps like image acquisition, image preprocessing, feature extraction and classification. The texture features of soil images are extracted using the low pass filter, Gabor filter and using color quantization technique. Mean amplitude, HSV histogram, Standard deviation are taken as the statistical parameters.

**OBJECTIVES**

* The main task is to automate the procedure. With the emerging of image processing and machine learning we can efficiently classify the soil sample in to groups which it belong to.
* Thus, soil characteristics identification and classification is very much important. Identification of the soil type helps to avoid agricultural product quantity loss.

# INTRODUCTION

Soil is the term which has different meaning for different people: for a geologist it represents the products of past surface processes. To a penologist it represents physical and chemical processes occurring currently. For an engineer soil is the solid thing up on which foundation for houses, factories, building, roads, etc can be built. Soil study means the knowing of externally identifiable patterns seen on soil. Grouping of soil is particularly basic for reasonable agricultural business. Recognizing the characteristics of soil is the key feature to reduce the product quantity losses. Knowing the type of soil is very useful for cultivation, construction..etc. As far as plant is concerned plantation according to the soil characteristics is very much important for its success.

The nature of soil is influenced by many factor, some of them are power of hydrogen (PH), Exchangeable sodium percentage, moisture content…etc. depending on their amount in soil they show different characteristics and that varies for different region. In preparation manual segmentation and classification method is monitored. This is time consuming, requires efficient people and expensive also. The main task is to automate the procedure. With the emerging of image processing and machine learning we can efficiently classify the soil sample in to groups which it belong to. Paper describes classification of the found segments using Machine Learning (ML) method Support Vector Machines (SVM).

**LITERATURE SURVEY**

**TITLE:** An algorithm for clustering and classification of series data with constraint of contiguity.

**AUTHOR:** B. Bhattacharya, and D.P. Solomatine

**YEAR: 2003**

**ABSTRACT:** Clustering and classification of series-based data is an important issue in a number of engineering problems, in particular in geotechnics. In such problems intervals of measured seria (signals) are. to be attributed a class so that the constraint of contiguity have to be considered and standard classification methods could be inadequate. Classification in this case needs involvement of an expert who observes the magnitude and trends of the signals in addition to any a priori information that might be available. In this paper an approach for automating this classification procedure is presented. Firstly, a segmentation algorithm is applied to segment the measured signals. Secondly, the salient features of these segments are extracted using boundary energy method. Based on the measured data and extracted features classifiers to assign class values to the segments were built; they employ decision trees and artificial neural networks. The algorithm was tested in a case-study for classifying sub-surface soil using measured data from cone penetration testing and satisfactory results were obtained.

**DESCRIPTION:** B.Bhattacharya..etc al.,uses the concept of segmentation , feature extraction and classification. The signals which are measured segmented using segmentation algorithms. Boundary energy method is used for extracting features from the input data. Depending on these features classifiers such as ANN, SVM and decision trees are employed and satisfactory results are obtained.

**TITLE:** Statistical to fuzzy approach toward CPT soil classification

**AUTHOR:** Z. Zhang, and M. T. Tumay,

**YEAR: 2005**

**ABSTRACT:** A soil engineering classification derived from the cone penetration test (CPT) involves the uncertainty of correlation between soil composition and soil mechanical behavior. This uncertainty results in overlaps of different soil types in currently used CPT soil classification systems. Accordingly, two statistical soil classification criteria, region estimation and point estimation, are suggested to address this problem. Further, a new fuzzy subset approach is introduced to develop a truly independent CPT soil engineering classification, and to establish a transition between the new fuzzy approach and conventional soil classifications by utilizing local site- and project-specific calibrations. CPT results conducted at the National Geotechnical Experimentation Site at Texas A&M University are used to demonstrate this new CPT soil engineering classification methodology.

**DESCRIPTION:**  Zhongjie Zhang..etc..al., the paper explains that there will be an uncertainty between the soil composition correlation and mechanical behavior of soil while deriving from CPT. This uncertainty leads to overlapping of different soil classes. The existing method available for this is point and region estimation. The author introduces a new fuzzy approach here that is independent of CPT.

**TITLE:** An analysis technique for biological shape.

**AUTHOR:** I.T. Young, and T.W. Calvert,

**ABSTRACT:** The motivation for a syntactic theory of shape was a belief that computer shape analysis could mimic human recognition procedures. The results described here suggest that heuristic techniques can be used to decompose and analyze complex objects. The derived decomposition algorithm gives quite good results when compared with human lobe determinations. In addition, it seems to follow closely the actual procedures used by observers in classifying shapes. The list structure output of the program clearly describes each shape element and greatly facilitates further analysis. Extensions of this work should be directed toward the development of more specific "rules" for shape decomposition, and more complete orientation information in the output list structure

**DESCRIPTION:** I.T. Young..etc..al, introduces a concept for analyzing the technique for biological shape based on bending energy. It finds out the amount of work used for typical biological shape in addition to this it explains sampling theorem for contours which are connected and closed and a fast algorithm for calculating the bending energy.

**TITLE:** Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations,

**AUTHOR:** I.H. Witten, and E. Frank,

**ABSTRACT:** We are overwhelmed with data. The amount of data in the world, in our lives, seems to go on and on increasing—and there’s no end in sight. Omnipresent personal computers make it too easy to save things that previously we would have trashed. Inexpensive multigigabyte disks make it too easy to postpone decisions about what to do with all this stuff—we simply buy another disk and keep it all. Ubiquitous electronics record our decisions, our choices in the supermarket, our financial habits, our comings and goings. We swipe our way through the world, every swipe a record in a database. The World Wide Web overwhelms us with information; meanwhile, every choice we make is recorded. And all these are just personal choices: they have countless counterparts in the world of commerce and industry. We would all testify to the growing gap between the *generation* of data and our *understanding* of it. As the volume of data increases, inexorably, the proportion of it that people understand decreases, alarmingly. Lying hidden in all this data is information, potentially useful information, that is rarely made explicit or taken advantage of.

**EXISTING SYSTEM**

A cone penetration test (CPT) is one among the popular soil investigation method. It is used for modeling the sub-surface soil and for a little depth information gathering from collected soil samples. A constraint limits the solutions available. The paper gives a survey on constrained classification. the paper handles with various algorithm on classification, properties of classes on division and the topologies of decision tree diagrams.

The paper uses few parameters for representing complex geological models using principle component analysis (PCA). Normal PCA works by performing multiplication using basis matrix and makes high dimensional model. Here optimization is used for mapping (O-PCA) which have non-Gaussian characteristics and enhance the features. Thus it is used for reducing gradient based approaches and to improve the matching process.

**EXISTING SYSTEM DRAWBACKS**

* In preparation manual segmentation and classification method is monitored. This is time consuming, requires efficient people and expensive also.
* Less effective feature extraction methods used in existing systems.
* The existing systems are less efficient and time-consuming.

**PROPOSED SYSTEM**

The main task is to automate the procedure. With the emerging of image processing and machine learning we can efficiently classify the soil sample in to groups which it belong to. This paper describes classification of the found segments using Machine Learning (ML) method Support

Vector Machines (SVM). This paper explains support vector machine based classification of the soil types. Soil classification includes steps like image acquisition, image preprocessing, feature extraction and classification. The texture features of soil images are extracted using the low pass filter, Gabor filter and using color quantization technique. Mean amplitude, HSV histogram, Standard deviation are taken as the statistical parameters. With more data and soil science domain-specific tricks, the potential for applying machine learning to soil property prediction would surely be maximized.

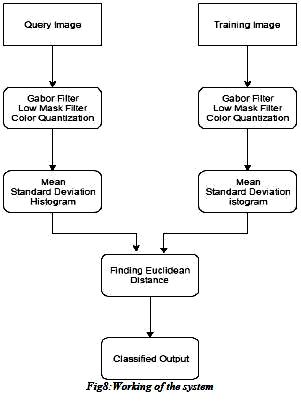
In order to classify the soil there are certain steps to be followed

1. Apply the transformation (low mask filter, color quantization , histogram ) to the original image

2. Use statistical measures to analyses the color ,texture, shape

3. Finding the distance with Euclidean distance formula

**PROPOSED SYSTEM BLOCK DIAGRAM**

****

**FLOW CHART:**

Image Pre-processing

Query Image

Image Filtering & Enhance Contrast

Feature Extraction

Classifications

Clay

Median Filter

(Gabor Filter , low pass filter , color Momentts)

Soil images

Feature Parameters

(Mean, Standard deviation , histogram )

SVM

Silty Sand

sandy clay

Peat

Humus clay

clayey sand

Clayey Peat

**PROPOSED METHODOLOGIES**

1. **Support Vector Machine(SVM)**

Support vector machine (SVM) in machine learning are supervised models associated with respect to learning algorithms. These are mainly used for analyzing the data for regression and classification. For a set of training examples it belongs to either one of the two categories, a support vector machine algorithm for training generates a model which tells the new thing falls in to which category by a non-probabilistic binary classifier. SVM model is the example on depiction of points in space which is mapped. Thus, the data of different types are separated by as wide as possible.

1. **Transformation**

1) Color quantization

The objective of color quantization or color image quantization is to make new image visually similar to that of the original image. Thus, reduces the distinct colors used in original image.

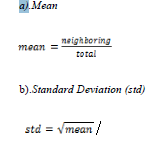
2) Low Pass Filter

A low-pass filter passes frequency below the cutoff frequency and attenuates the higher frequency. The attenuated frequency depends on the filter design. In audio applications this also called as high-cut filter or treble –cut filter.

3) Gabor Filter

Gabor filter also known as linear filter used as an edge detector. For the extraction of feature from an image Gabor filter with different frequency are useful. In image processing a 2-D Gabor filter is used for feature extraction specially while doing segmentation and analyzing texture.

C. **Statistical Parameters**



**PROPOSED SYSTEM ADVANTAGES**

* Our proposed method shows better performance compared to existing.
* With more data and soil science domain-specific tricks, the potential for applying machine learning to soil property prediction would
* From a set of seven class the classifier have to identify the appropriate class in this area.
* The classifications of non-sandy soils are better classified with SVM (through WEKA).

**SOFTWARE REQUIREMENTS**

* MATLAB 7.14 Version R2012

**MATLAB**

The MATLAB high-performance language for technical computing integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation.

* Data Exploration ,Acquisition ,Analyzing &Visualization
* Engg drawing and Scientific graphics
* Analyzing of algorithmic designing and development
* Mathematical functions and Computational functions
* Simulating problems prototyping and modeling
* Application development programming using GUI building environment.

Using MATLAB, you can solve technical computing problems faster than with traditional programming languages, such as C, C++, and Fortran.