# Performance of SVM Classifier For Image Based Soil Classification

Srunitha.k
Dept of Computer Science and Engineering
Amrita School of Engineering, Coimbatore
Amrita vishwa vidyapeedham
Amrita University, India
srunithagknair@gmail.com

Dr.S.Padmavathi
Dept of Computer Science and Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham
Amrita University, India
s_padmavathi@cb.amrita.edu

***Abstract*— 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 pre-processing, 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.**

***Keywords—SVM; HSV; Gabor filter;***

## I. 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. Soils may be described in different ways by different people for their different purposes. 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. It is crucial for countries that export several agricultural commodities. A classification for engineering purposes should be based mainly on mechanical properties, e.g. permeability, stiffness, strength. The class to which a soil belongs can be used in its description. 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. Soil type of a particular geographical area is analyzed by collecting samples of soils and classifying them in to different type using different methodologies. 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. This paper describes classification of the found segments using Machine Learning (ML) method Support Vector Machines (SVM).

The paper contains the following; Section 1 is the brief introduction about the soil classification need. Section II describes the literature review. Section III is the methods used in details. Section IV is the dataset details. Section V is about the proposed algorithm explanation. Section VI contains the experimental results and Finally Conclusions are summarized in section VII.

## II. LITERATURE SURVEY

B.Bhattacharya..etc al.,[3] 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.

A cone penetration test (CPT) is one among the popular soil investigation method [4]. It is used for modelling the sub-surface soil and for a little depth information gathering from collected soil samples.

A constraint limits the solutions available. The paper[6] 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 [8].

978-1-5090-4620-1/16/$31.00 ©2016 IEEE

A Comprehensive Foundation on neural networks [7] provides an inclusive overview on the neural network applications. An overview in knowledge illustration along with in what way they are used in artificial intelligence (AI) is given. The vapnik-chervonenkis (VC) dimension is explained here with respect to the training samples that a machine can absorb without errors. This contains information about the least mean square error, back propagation …etc.

Fuzzy measure is used as base for aggregating the correlation between relative densities and CPT. Here three levels high, low, medium sands are selected by friction ratio. Based on the differences between these levels the compressibility is measured. Based on fuzzy c-means and integrals the correlation density is measured. Obtained results are compaired here [10].

Decoding of Bose Chaudhuri Hocquenghem codes (BCH) is explained using multiclass SVM. Normal algorithm the decoding is fixed regardless of the SNR environment. Thaw there is no local minima and outlier robust SVM shows a great capacity in decoding the BCH codes [1]. ABDF is feature enriched and helpful and provides GUI for examine the huge data. [2]

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 [22].

I.T. Young..etc..al, [21] 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.

R. Webster ..etc al., introduces two method in Optimally partitioning soil transects[19]. One method is by using a window called split moving window (SMW) and the other one is maximum level variance (MLV). The transect is examined through the SMW and the values on the other side of the mid-point is compared. MLV do examines each and every possible regions thus minimizing the within square variance. A comparisons and calculation of these methods is summarized over here.

## III. METHODS USED

### A. *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. The new data are mapped and categorized according to which part of the group they fall on. Support Vector Regression: Developed by Vapnik (1998).

Support vector machine through nonlinear mapping maps, the input vector X to high-dimensional feature space (F). Here only one hyperplane which is optimal is created. Let the training set D he {(xi, yi)xI, with input xi and yi = {* 1). The SVM format maps x to z =@@)E F . When the data is linearly separable in F , the SVM constructs a hyperplane w =z+b for which separation between the positive and negative examples is maximized.

It can be shown that w= xa,y,z, , , where a =(al, ..., aN) can he found by solving the following quadratic programming problem: min-a Qa-1 a T N ;=I T (1) IT 2 T subject to a20 and a y=O. where, y=(~~,....y~)~ and Q has entries yiyjzrzj = yiyjK(x;,xj), where K(xi,xj) is called a kernel. When the training set is not separable in F , the SVM algorithm introduces non-negative slack variable 6, 2 0. The multivariate problem becomes subject to yi(wrzi +b)21-4; . C is a regularization parameter controlling the tradeoff between model complexities and training error. The xi for which ai # 0 are defined as the support vectors, since they determine the optimal hyperplane, the hyperplane with maximal margin. The support vectors geometrically fall closest to the optimal hyperplane.

### B. *Basic segmentation method*

A segmentation process splits region of interest from that of non-interest regions. A two class classifier is essential for classifying pixels in feature space considering segmentation as a two class problem. The method of segmentation is explained below.

Training

1) As training data one or a few images having objects are considered. Traditional segmentation or by manually foreground and background regions are splited. Pixels in objects are marked using I and I- which produces RGB color histogram. Color values are also marked.
2) Prepare for SVM the training data
(xi,yi) , + + +l,if Xi€[ . xi is -1,if X,EI. +- Xi€[ VI , yi= a color vector.

### C. *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. A two-dimensional Gabor filters in the discrete domain,

$$G_c[i,j] = Be^{-\frac{(i^2+j^2)}{2\sigma^2}} \cos(2\pi f(i\cos\theta + j\sin\theta))$$

$$G_s[i,j] = Ce^{-\frac{(i^2+j^2)}{2\sigma^2}} \sin(2\pi f(i\cos\theta + j\sin\theta))$$

In image processing a 2-D Gabor filter is used for feature extraction specially while doing segmentation and analyzing texture. The term $f$ corresponds to frequency in the texture. The orientation of texture in a particular direction can be obtained by varying $\theta$ value. Analyzed image region size and basis support can be changed by varying $\sigma$ component

D. *Statistical Parameters*

*a).Mean*

$$mean = \frac{neighboring}{total}$$

b).*Standard Deviation (std)*

$$std = \sqrt{mean}$$

## IV. DATASET

The dataset consisted of a collection of 175 soil sample measures. The dataset contain Silty Sand, Sandy Clay, Peat, Humus Clay, Clayey Sand, Clayey Peat, Clay.

Some of the dataset images is as follows:

*Fig 1: Clay*

*Fig 2: Clayey Peat*

*Fig 3: Clay Sand*

*Fig 4: Humus Clay*


*Fig 5: Peat*

*Fig 6: Sandy Clay*

*Fig 7: Silt Sand*

## V. STEPS TO BE FOLLOWED

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

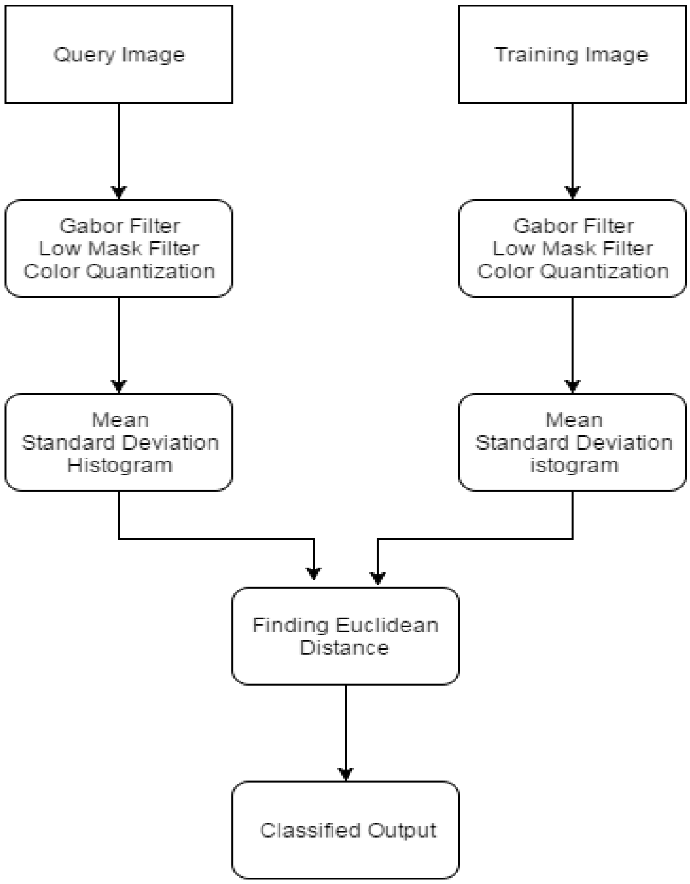


*Fig8:Working of the system*

## VI. EXPERIMENTL RESULTS

The system recognizes and classifies different type of soils like the result as, shown in figure.

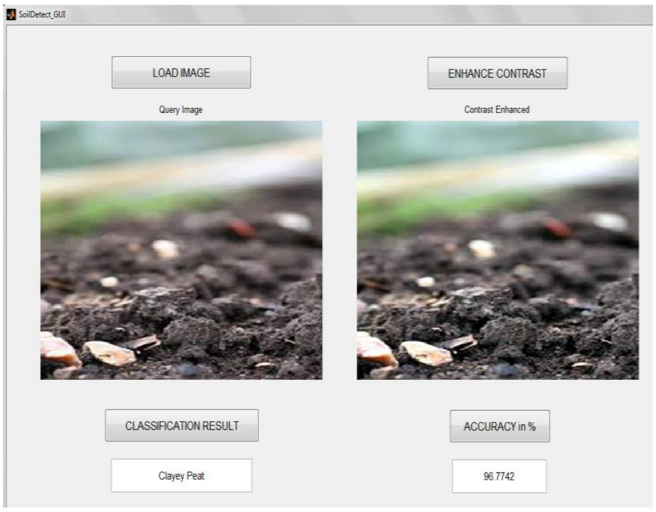


*Fig9:result*

1. Binary classification: binary classifier (Table:1) is used to govern whether the soil type is sandy or not. Support vector machine (SVM) do the classification of the non-sandy soil. The soil types are better classified here (with WEKA). Majority of the misclassified objects are relayed near to the segment line. Near the segment boundary Measurements spotted as often noisy and thus can be decided that the enactment of the classifiers was excellent.

| **Soil class** | **Percentage of correctly classified segments** | **Percentage of correctly classified instances** |
|---|---|---|
| Sandy soil | 100 | 96.6 |
| Non-sandy soil | 100 | 99.5 |
| Total | 100 | 97.8 |

*Table 1: Binary classification*

2. Three-class classification (Table 2): Only primary class soils like clay, peat and sand are classified here by the classifier. Because of the very high overlapping of clayey soil classes with that of other two classes, three class classifications is much more difficult comparing with binary class. The performance of classifier in terms of SVM (with WEKA) is comparable and gives better result. SVM classifier is better for clayey soil.

| **Soil class** | **Percentage of correctly classified segments** | **Percentage of correctly classified instances** |
|---|---|---|
| Sand | 100 | 98.5 |
| clay | 100 | 98.5 |
| Peat | 100 | 58.7 |
| Total | 100 | 90..7 |

*Table 2: Three class-classification*

3. Seven-class classification: From a set of seven class the classifier have to identify the appropriate class in this area. The SVM-based classifiers gave poor results (Table 3) these results are, however, preliminary since no optimization of the built SVMs (of regularizations constants and the kernels) was undertaken

| **Soil class** | **Percentage of correctly classified segments** | **Percentage of correctly classified instances** |
|---|---|---|
| Silty sand | 100 | 85.1 |
| Clayey sand | 100 | 54.2 |
| Sandy clay | 50 | 69.4 |
| Clay | 100 | 65.0 |
| Humus clay | 80 | 77.0 |
| Clayey peat | 28.6 | 49.7 |
| Peat | 50 | 65.4 |
| Total | 60.9 | 74.4 |

*Table 3: Seven class- classification*

## VII. CONCLUSION

The classifications of non-sandy soils are better classified with SVM (through WEKA). Almost all misclassified objects are relayed near to the segment line. Near the segment boundary Measurements spotted as often noisy and thus can be decided that the enactment of classifiers was excellent. With more data and soil science domain-specific tricks, the potential for applying machine learning to soil property prediction would surely be maximized. It is able to achieve a 95% accuracy rate for classifying.

## References


[1]. V. Sudharsan and B. Yamuna "Support Vector Machine based Decoding Algorithm for BCH Codes" Journal of Telecommunication and Information Technology 2016.

[2]. Unmesha Sreeveni.U .B, Shiju Sathyadevan "ADBF Integratable Machine Learning Algorithms –Map reduce Implementation" Second International Symposium on computer vision and the Internet(VisionNet'15).

[3]. B. Bhattacharya, and D.P. Solomatine "An algorithm for clustering and classification of series data with constraint of contiguity", Proc. 3T"d nt. Conf: on Hybrid and Intelligent Systems, Melboume, Australia, 2003, pp. 489-498.

[4]. A.Coerts, Analysis of Static Cone Penetration Test Data for Subsurface Modelling - A Methodology (PhD Thesis),Utrecht University, The Netherlands, 1996.

[5]. L.F. Costa, and R.M. Cesar, Shape Analysis and Classification: Theory and Practice, Boca Raton, Florida: CRC Press, 2001.

[6]. Gordon, A.D. "A survey of constrained classification", Computational Statistics & Data Analysis, vol. 21, pp. 17-29, 1996.

[7]. S. Haykin, Neural Networks: A Comprehensive Foundation, New Jersey: Prentice Hall, 1999.

[8]. D.M. Hawkins, and D.F. Merriam, "Optimal zonation of digitized sequential data", Mathematical Geology, vol. 5, pp. 389-395, 1973.

[9]. G.P. Huijzer, Quantitative Penetrostratigraphic Classification (PhDThesis), Free University of Amsterdam, The Netherlands, 1992.

[10]. C.H. Juang, X.H. Huang, R.D. Holtz, and J.W. Chen, "Determining relative density of sands from CPT using fuzzy sets", J. of Geotechnical Engineering, vol. 122(1), pp. 1-6, 1996.

[11]. M.G. Kerzner, Image Processing in Well Log Analysis, Dordrecht,The Netherlands: Reidel Pub., 1986.

[12]. J. K. Kumar, M. Konno, and N. Yasuda,"Sub surface soil-geology interpolation using fizzy neural network", J. of Geotechnical and Geoenvironmental Engineering, ASCE, vol. 126(7), pp. 632-639, 2000.

[13]. Neural Machine, http:/ivww.data-machine.com/, 28.1.2005.

[14]. Neuro Soluions, http://www.nd.com/, 28.1.2005.

[15]. RHUL , Computer Learning Research Centre, Royal Holloway University of London, (http://www.clrc.rhul.ac.uk, 26/1/2005)

[16]. L.J. van Vliet, and P.W. Verbeeck, "Curvature and bending energy in digitised 2D and 3D images", in: K.A. Hogda, B. Braathen and K.Heia (Eds), Proc. 8" Scandinavian Confon Image Analysis, Norway,1993, vol. 2, pp. 1403-1410.

[17]. K. Wagstaff, Intelligent Clustering with Instance-Level Constraints(PhD thesis), Cornell University, USA, 2002.

[18]. R. Webster, "Optimally partitioning soil transects", Journal of Soil Science, vol. 29, pp. 388402, 1978.

[19]. H.J.T. Weerts, Complex Confining Layers, Utrecht University, The Netherlands, 1996.

[20]. I.H. Witten, and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufimann, 2000.

[21]. I.T. Young, and T.W. Calvert, "An analysis technique for biological shape", Information and Control, vol. 25, pp 357-370, 1974.

[22]. Z. Zhang, and M. T. Tumay, "Statistical to fuzzy approach toward CPT soil classification", J of Geotechnical and Geo environmental Engineering, vol. 125(3), pp. 179-186, 1999.