Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE xxx (xxxx) xxx

journal homepage: www.elsevier.com/locate/inpa

Soil texture classification using multi class support vector machine

Utpal Barman [a,*], *Ridip Dev Choudhury* [b]

[a] *Department of IT, Gauhati University, Guwahati, Assam, India*
[b] *Department of Computer Science, IDOL, Gauhati University, Guwahati, Assam, India*

ARTICLE INFO

Article history:
Received 25 April 2018
Received in revised form
20 July 2019
Accepted 4 August 2019
Available online xxxx

Keywords:
Image processing
Soil
SVM
Classification

ABSTRACT

The objective of this study is to process the soil images to generate a digital soil classification system for rural farmers at low cost. Soil texture is the main factor to be considered before doing cultivation. It affects the crop selection and regulates the water transmission property. The conventional hydrometer method determines the percentage of sand, silt, and clay present in a soil sample. This method is very cost and time-consuming process. In this approach, we collect 50 soil samples from the different region of west Guwahati, Assam, India. The samples are photographed under a constant light condition using an Android mobile of 13 MP cameras. The fraction of sand, silt, and clay of the soil samples are determined using the hydrometer test. The result of the hydrometer test is processed with the United State Department of Agriculture soil classification triangle for the final soil classification. Soil images are processed through the different stages like pre-processing of soil images for image enhancement, extracting the region of interest for segmentation and the texture analysis for feature vector. The feature vector is calculated from the Hue, Saturation, and Value (HSV) histogram, color moments, color auto Correlogram, Gabor wavelets, and discrete wavelet transform. Finally, Support Vector Machine classifier is used to classify the soil images using linear kernel. The proposed method gives an average of 91.37% accuracy for all the soil samples and the result is nearly the same with the United State Department of Agriculture soil classification.

1. Introduction

The physical and chemical properties of the soil always play an important role in precision and smart farming. Soil texture is the main physical property of the soil and plays an important role in farming [39]. It defines the fraction of sand, silt, and clay present in a soil sample. Other related agriculture cultivating properties such as water content of the soil, plant growth, and crop selection relate to these three end members. Most of the rural farmers are unknown about the texture and characteristics of the soil. Sandy soil is low in the water holding capacity and organic matter. The water holding capacity of silt soil is more as compared to sandy soil [25]. Clay soil shows the high water holding capacity, high plasticity and thickness whereas sandy soils are conspicuous by the absence of these properties. The water holding capacity of sandy loam is low as compared to loam and silty clay soil

* Corresponding author.
E-mail addresses: utpalbelsor@gmail.com (U. Barman), rdc@gauhati.ac.in (R.D. Choudhury).
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2019.08.001
2214-3173

[2]. It indirectly affects the plant growth, the weight of the plant and the percentage of chlorophyll. For example, chlorophyll contents of leaves are dropped in sandy loam as compared to silty clay soil [2].

Conventionally, the hydrometer method calculates the fraction of sand, silt, and clay of soil. Then, the USDA triangle method uses to classify the soil. All these methods are time and labour consuming processes. These are not appropriate approaches for rural farmers. In the field study, it is found that rural farmers do not have any knowledge about the soil texture. They are doing their farming without any proper soil testing and are unaware about the selection of soil and seed. It indirectly affects the overall growth of the plant. To overcome this problem, researches are done to develop the computer system and it efficiently classifies the soil texture using digital image processing.

Most of the work of soil texture classification is in three classes or multi-class. Among the three classifications, Zhang et al. [4] classified sand, silt, and clay soil using wavelet transform and maximum likelihood method. They incorporated fisher's linear discrimination analysis (FLDA) for feature dimension reduction and optimization. Sun et al. [1] forwarded a wavelet frame based method to classify the three classes of soil texture and achieved 91% accuracy for the classification. They achieved the same accuracy using Gabor wavelets [1]. Bhattacharya and Solomatine [8] introduced the decision tree, artificial neural network and support vector machine classifier to classify the soil texture on the segmented soil samples and achieved 90.7% accuracy for support vector machine. Zaho et al. [26] presented a model to predict the soil texture of high-resolution soil images. They applied an artificial neural network to predict the sand and clay contents of the soil in the collected soil maps. The neural network model was trained by the Levenberg-Marquardt optimization algorithm and Resilient back propagation algorithm but achieved more accuracy in the Levenberg-Marquardt optimization algorithm. The accuracy for Clay was 88% whereas the accuracy for Sand was 81%. Wu et al. [27] classified the clay, loam and sandy soil of the Yangtze River, Southwest China. They compared the performance of different machine learning approaches for soil classification such as support vector machine, artificial neural network, and classification tree. The support vector machine with the polynomial kernel gave a better result as compared to the other two one.

Among the multi-classification, Zhang et al. [5] introduced a discrete wavelet transform and linear discriminate analysis method for soil texture classification. They mixed the hyper spectral signals of sand, silt, and clay soil and classified using the maximum likelihood classifier. Later, they applied the Hidden Markov model for soil texture classification and achieved more accuracy as compared to maximum likelihood classifier [6]. Bhattacharya and Solomatine [8] applied decision tree, artificial neural network and support vector machine classifier to classify the soil texture from the segmented soil sample. They achieved an average of 70.4% accuracy for the classification and found that the seven support vector machine classifier gave the best result for silty sand and clay sand. Artificial neural network and decision tree model gave the best result for humus clay, peat, and the clay peat soil respectively. Shenbagavalli et al. [7] presented a soil texture classification method using mask convolution. They computed mean, abs mean, standard deviation, skewness, and kurtosis of the soil image and prepared as a feature vector. Chung et al. [9,36] introduced a soil texture classification model using RGB color. The RGB values are regressed with the actual laboratory values and got the coefficient of best fit as 0.96. Vibhute et al. [22] applied the radial base kernel of the SVM classifier on hyperspectral remote sensing soil data and classified five types of soil. The average accuracy of the model was 71.18% with the Kappa = 0.57. Breul and Gourves [3] presented a method of soil characterization using the third order moment of the input image. They recorded soil images using an endoscope.

It is come to know that soil texture property is the most important property for plant growth. During the literature survey, it is found that the researchers did valuable works for soil texture classification. However, most of the works are not in the aspect of a rural farmer as they collected their soil images using sensor or used hyper spectral soil images [4,8,22]. Sometimes the researchers collected the soil samples using a digital camera [9]. In this paper, 50 soil samples are photographed using an Android mobile device with 13 MP cameras. Capturing soil image using a Smartphone is the main challenging task for soil classification because the natural light can directly affect the quality of the soil image. Measuring all these factors, we developed our system in such a way that a rural farmer can easily capture a soil sample and use the system for soil classification.

2. Material and methods

2.1. *Soil sample collection*

The proposed soil classification system contains 50 soil samples. Samples are collected from 10 paddy fields of west Guwahati under the supervision of the Department of Civil Engineering, Girijananda Chowdhury Institute of Management and Technology (GIMT), Guwahati, Assam, India. The sample collection site lies in between 26.1445° N latitude and 91.7362° E longitude. Five samples are collected from each field with a separation of 200 m. Soils are dragged in a depth of 6 in. from the top-level of the paddy field and kept in a white paper for image acquisition. Hydrometer test is used to calculate the fraction of sand, silt, and clay in the soil and determined the texture of the soil using USDA triangle. In Fig. 1, the area under the black circle represents the site of sample collection.

2.2. *Image acquisition*

Image acquisition is the process of capturing soil images from the samples. A Xiaomi Redmi 3S smartphone with the marshmallow version of Android is used to photograph the soil images. The Smartphone camera consists of a 13-megapixel charged-coupled device. During the time of image capture, all the camera properties are kept in default settings such as F-stop = f/2, exposure time = 1/60 S, ISO speed = ISO 125,

Please cite this article as: U. Barman and R. D. Choudhury, Soil texture classification using multi class support vector machine, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2019.08.001

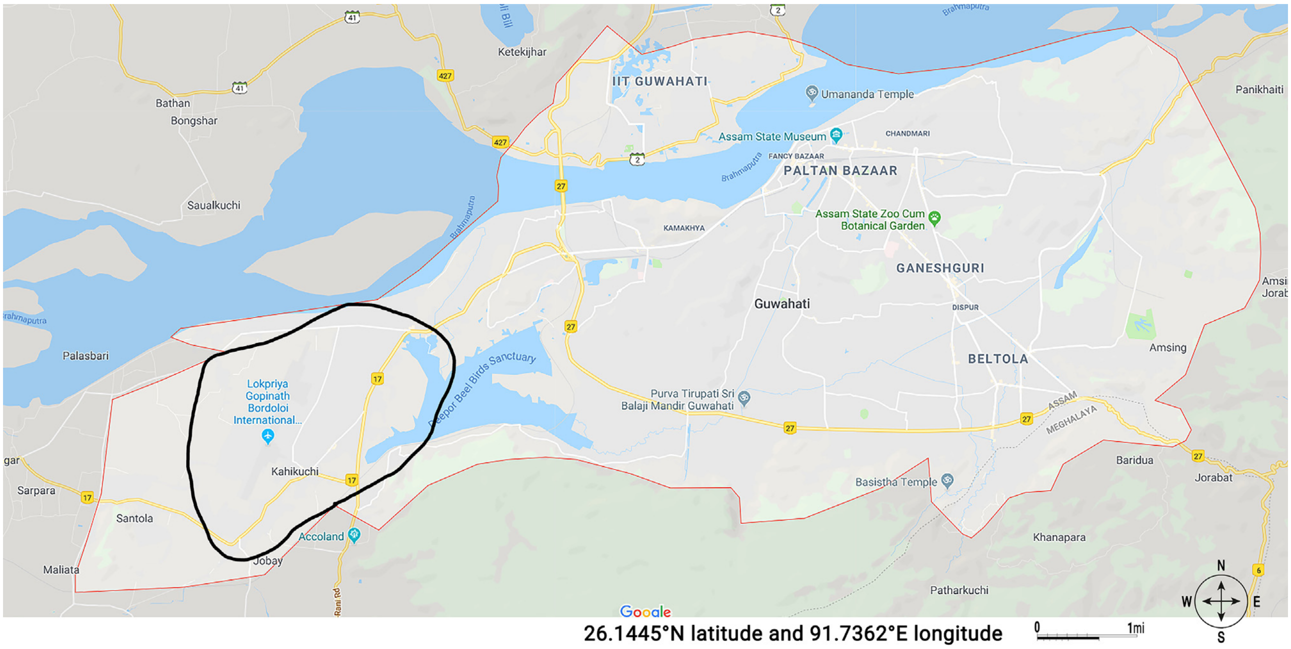

Fig. 1 – Area for Sample Collection [25].

focal length = 4 mm. Fig. 2a shows the procedure of the image acquisition. Soil samples are photographed in a closed room with only one curtain less window. Light enters into the room through the window. The effect of natural light is less in this process and is also very cost effective. In order to avoid the other environmental light condition, soil images are photographed from 10 in. ahead from the soil in a straight position as shown in Fig. 2a. All the soil images are kept as per field starting with a number from 1 to 5. Five samples are photographed from each field and a total of 50 soil samples are found using an Android Smartphone. Fig. 2b shows the samples of the captured soil images.

2.3. *Soil texture determination in laboratory*

In this section, the procedure of soil texture classification using the hydrometer and USDA triangle methods are explained. Soil consists of sand, silt, and clay [10]. The fraction of sand, silt, and clay gives an idea of the soil texture and are calculated using the pipette or hydrometer method [10]. The final classification is done using USDA criteria. The proposed system chose the hydrometer method as the laboratory method of soil texture classification as it clearly gives the fraction of sand, silt, and clay present in a soil sample [11]. The other reason for choosing the hydrometer method is that

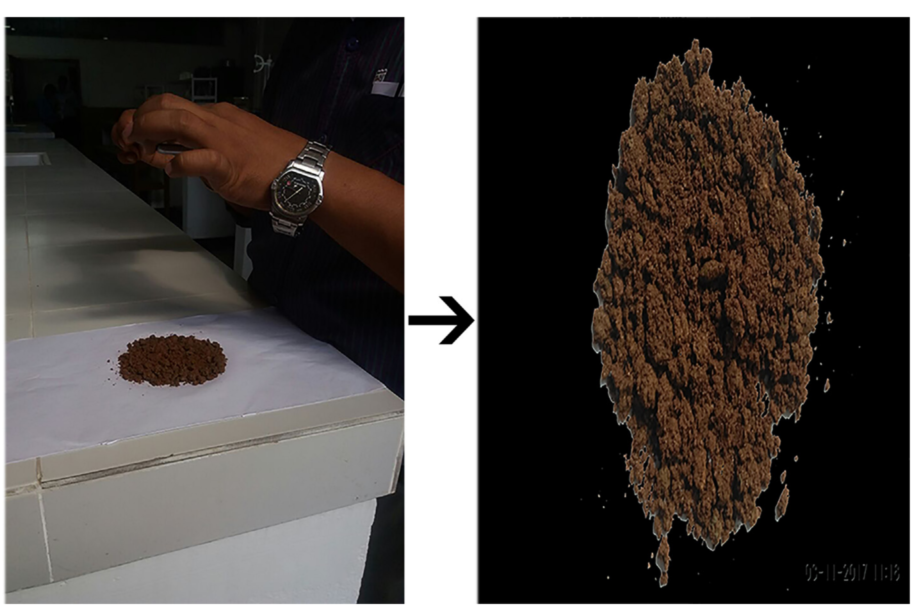

Fig. 2a – Soil Sample Collection and its ROI after Segmentation.

Please cite this article as: U. Barman and R. D. Choudhury, Soil texture classification using multi class support vector machine, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2019.08.001

Fig. 2b – Some of the samples of Collected Soil Image Database.

it is available in GIMT, Guwahati, and nearest to the sample collection site.

The hydrometer test is done in the civil engineering department of GIMT, Guwahati, Assam India. The hydrometer is calibrated by preparing a solution of 28-gram tablespoon sugar and 176 g of distilled water. The solution is prepared up to the calibration temperature of the hydrometer and measure the gravity of the solution to 1.048. After the calibration, all the soil samples are processed for hydrometer test. In the lab, 50 g of each soil sample within a limit of 2 mm coarse fragment are dispersed with sodium metaphosphate and then agitated. The entire process follows all the steps of the hydrometer test [11]. At last, the fraction of sand, clay, and silt is calculated using the following equations [11].

$$\%\,\text{Clay} = \text{Corrected hydrometer reading at 6hrs, 52 min} \times 100/\text{wt of sample} \quad \text{(i)}$$

$$\%\,\text{Silt} = \text{Corrected hydrometer readings at 40 sec} \times 100/\text{wt of sample} - \%\text{ of clay} \quad \text{(ii)}$$

$$\%\,\text{Sand} = 100\% - \%\text{Silt} - \%\text{Clay} \quad \text{(iii)}$$

The result of the hydrometer method is recorded in the percentage of sand, silt, and clay. The textures of the soil sample are determined using the United State Department of Agriculture (USDA) triangle method. For the implementation, a Matlab code is developed to classify the texture of the soil and the output of the USDA classification is represented in the excel sheet. The actual soil texture of sample 1 from the first paddy field is shown in Fig. 3 as per the USDA criteria.

In Fig. 3, the red point is located in the area of silt loam. Loam soil means the combination of sand, silt, and clay. Silt loam expresses all the fraction of three end members of the

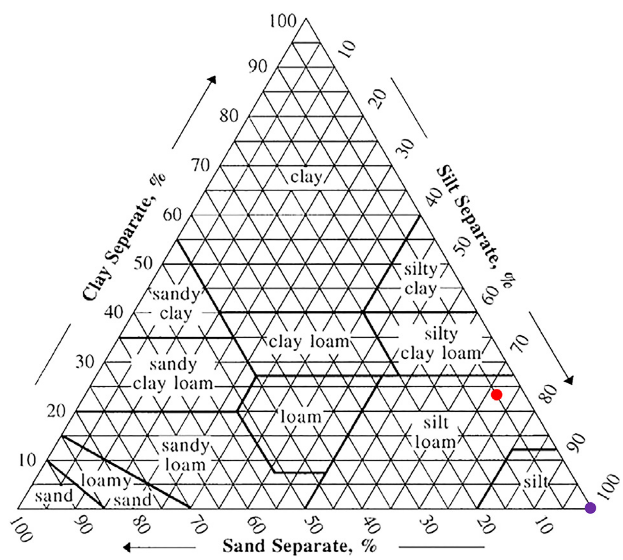

Fig. 3 – USDA Triangle Value for Sample 1. [18]

soil, but the fraction of silt is more as compared to the other two one.

In the hydrometer test for 50 soil samples, it is found that

(a) A total of 4 samples out of 50 samples are clay and these belong to paddy fields 3 and 9.
(b) A total of 7 samples out of 50 samples are clay loam and these belong to paddy fields 3, 5, 8 and 9.
(c) Only 2 samples out of 50 samples are fine sand and these belong to paddy field 2.
(d) Only 2 samples out of 50 samples are loam and these belong to paddy field 5

Please cite this article as: U. Barman and R. D. Choudhury, Soil texture classification using multi class support vector machine, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2019.08.001

Table 1 – Mean Fraction of Sand, Silt, and Clay of each soil Sample.

Mean % Sand	Mean % Silt	Mean % Clay	Texture (USDA)
23.51	15.38	61.12	Clay
31.87	36.39	31.72	Clay Loam
95.00	0.00	2.50	Find Sand
38.00	38.13	23.81	Loam
85.00	5.00	10.00	Loamy Fine Sand
85.00	3.00	12.00	Loamy Sand
54.64	6.40	39.08	Sandy Clay
61.52	12.58	26.19	Sandy clay loam
19.15	63.68	17.16	Silt Loam
20.00	40.00	40.00	Silty Clay
13.21	53.95	32.83	Silty Clay Loam

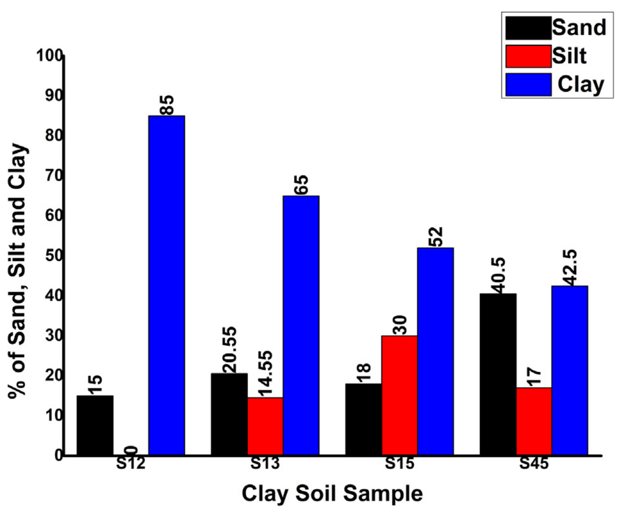

Fig. 4 – Fraction of Clay Soil Sample.

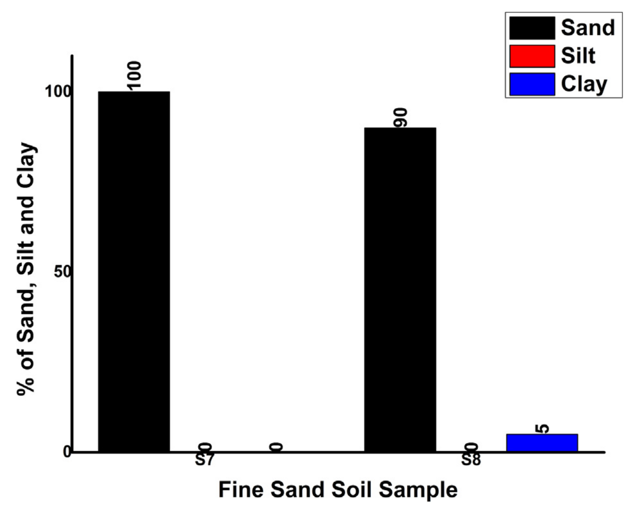

Fig. 6 – Fraction of Fine Sand Soil Sample.

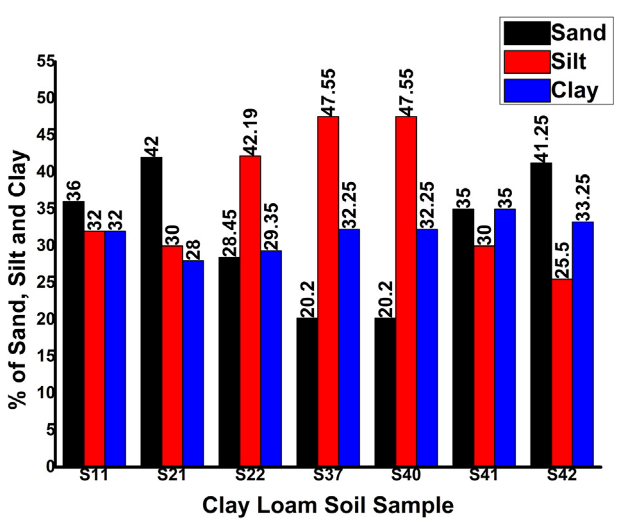

Fig. 5 – Fraction of Clay Loam Soil Sample.

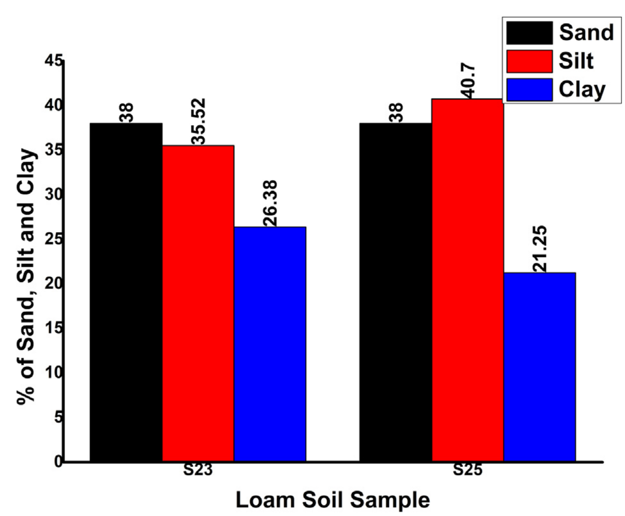

Fig. 7 – Fraction of Loam Soil Sample.

(e) Only 1 sample out of 50 samples is loamy fine sand and it belongs to paddy field 4.

(f) Only 1 sample out of 50 samples is loamy sand and it belongs to paddy field 6.

(g) A total of 3 samples out of 50 samples are sandy clay and these belong to paddy fields 4 and 9.

(h) A total of 10 samples out of 50 samples are sandy clay loam and these belong to paddy fields 2, 4, 6, 7 and 9.

Please cite this article as: U. Barman and R. D. Choudhury, Soil texture classification using multi class support vector machine, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2019.08.001

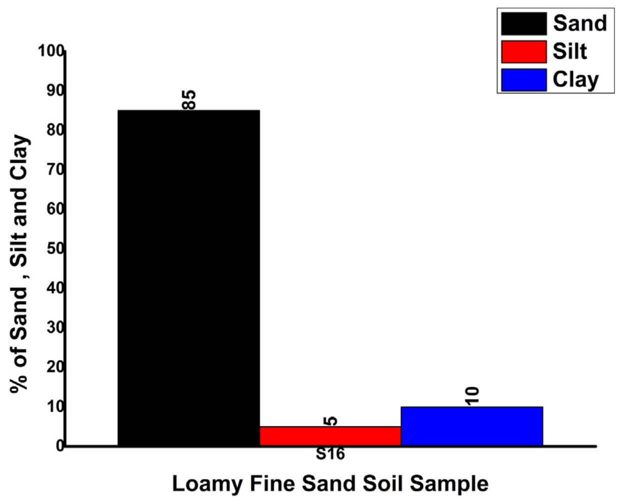

Fig. 8 – Fraction of Loamy Fine Sand Soil Sample.

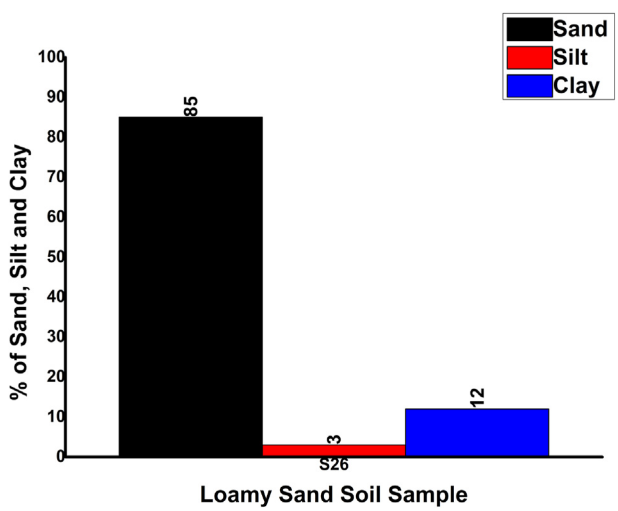

Fig. 9 – Fraction of Loamy Sand Soil Sample.

(i) A total of 5 samples out of 50 samples are sandy loam and these belong to paddy fields 2, 6 and 7.
(j) A total of 11 samples out of 50 samples are silt loam and these belong to paddy fields 1, 5 and 8.
(k) Only 1 sample out of 50 samples is silt clay and it belongs to paddy field 1.
(l) A total of 3 samples out of 50 samples are silt clay loam and these belong to paddy fields 3 and 8.

The mean fraction sand, silt and clay of each soil sample along with USDA classification is presented in Table 1.

Fig. 4 shows the graphical value of the fractions of clay soil for the samples 12, 13, 15, and 45.

Fig. 5 shows the graphical value of the fractions of clay loam soil for the samples 11, 21, 22, 37, 40, 41, and 42.

Fig. 6 shows the graphical value of the fractions of fine sand soil for the samples 7, and 8.

Fig. 7 shows the graphical value of the fraction of loam soil for the samples 23, and 25.

Fig. 8 shows the graphical value of the fraction of loamy fine sand soil for the sample 16.

Fig. 9 shows the graphical value of the fraction of loamy sand soil for the sample 26.

Fig. 10 shows the graphical value of the fraction of sandy clay soil for the samples 19, 20, and 43.

Fig. 11 shows the graphical value of the fraction of sandy clay loam soil for the samples 10, 17, 18, 27, 29, 31, 32, 33, 34, and 44.

Fig. 12 shows the graphical values for all the fraction of sandy loam for the samples 6, 9, 28, 30, and 35.

Fig. 13 shows the graphical value of the fraction of silt loam soil for the samples 1, 2, 3, 5, 24, 38, 46, 47, 48, 49, and 50.

Fig. 14 shows the graphical value of the fraction of silt clay soil for the sample 4.

Fig. 15 shows the graphical values of the fraction of silty clay loam soil for the sample 4.

After the hydrometer method, USDA classification gives 12 different types of soil classes including Clay, Clay Loam, Loam, Fine Sand, Sandy Clay loam, etc. The samples of each class are divided into training and testing set. It is presented

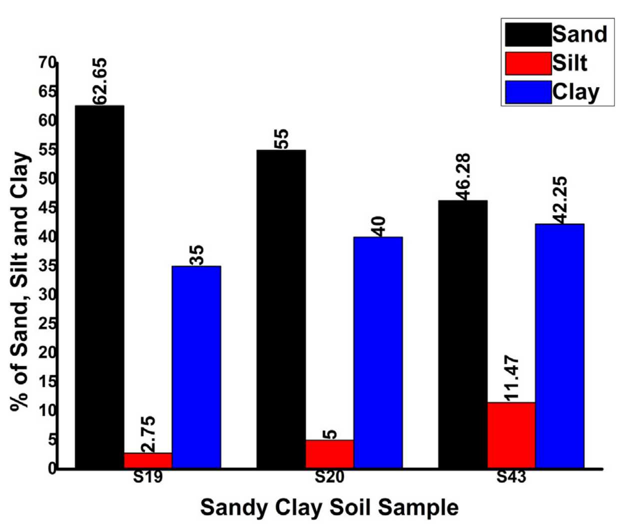

Fig. 10 – Fraction of Sandy Clay Soil Sample.

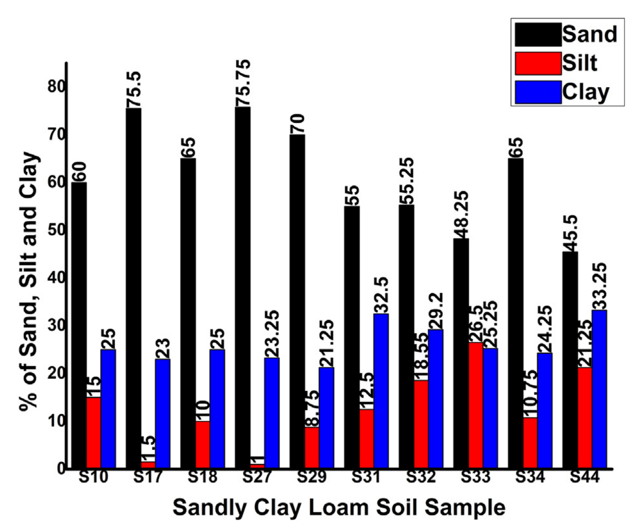

Fig. 11 – Fraction of Sandy Clay Loam Soil Sample.

Please cite this article as: U. Barman and R. D. Choudhury, Soil texture classification using multi class support vector machine, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2019.08.001

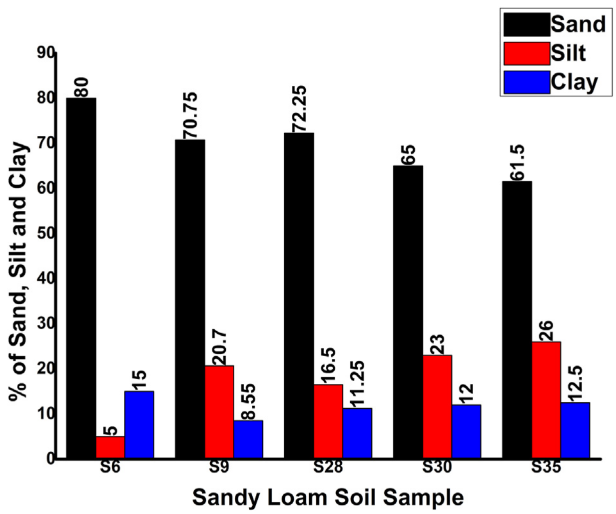

Fig. 12 – Fraction of Sandy Loam Soil Sample.

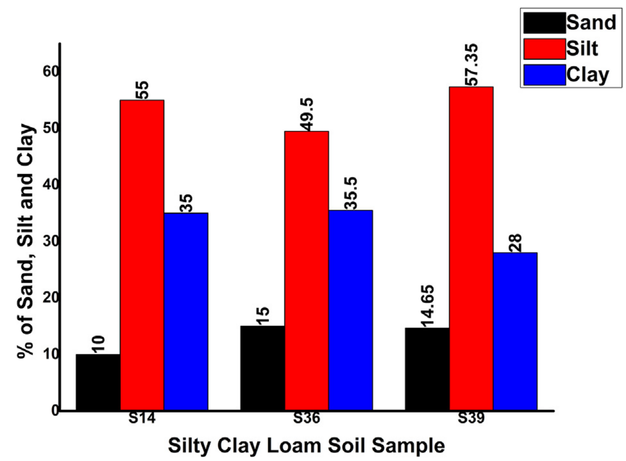

Fig. 15 – Fraction of Silty Clay Loam Soil Sample.

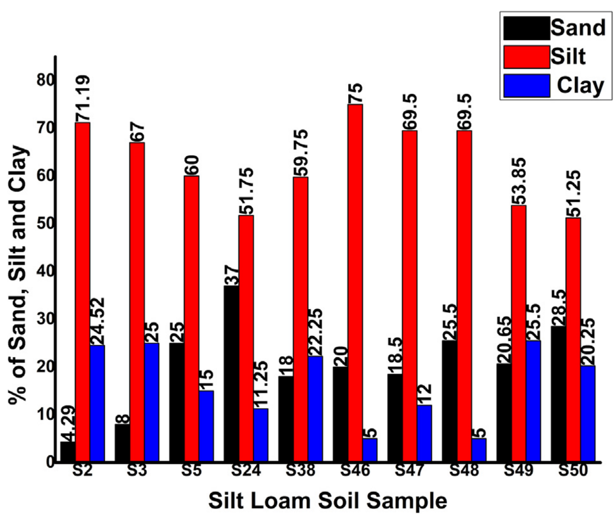

Fig. 13 – Fraction of Silt Loam Soil Sample.

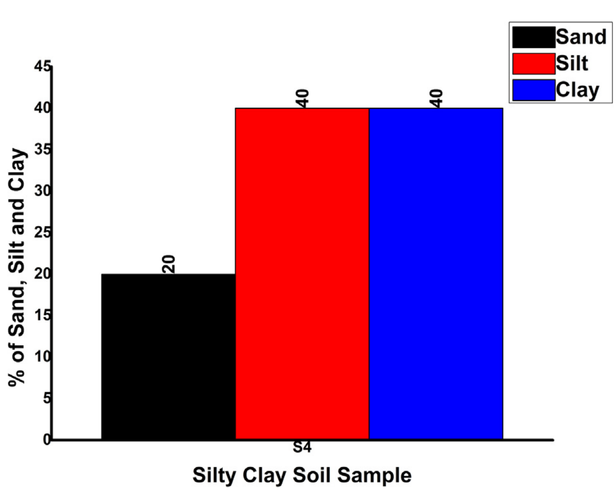

Fig. 14 – Fraction of Silt Clay Soil Sample.

Table 2 – Description of training and testing samples of soil image.

Soil class	Total Number of Samples	Training Image	Testing Images
Clay Loam	7	4	3
Sandy Clay Loam	10	5	5
Silt Loam	11	6	5
Clay	4	2	2
Fine Sand	2	1	1
Loam	2	1	1
Loam Fine sand	1	1	1
Loamy Sand	1	1	1
Sandy Clay	3	2	1
Sandy Loam	5	3	2
Silt Clay	1	1	1
Silt Clay Loam	3	2	1
Total	50	29	24

in Table 2. Among all the soil classes, 3 soil classes such as Fine sand, Loamy sand, and Silt Clay soil have only 1 image. So, the same image is considered for training and testing.

2.4. *Soil image Prepocesing, segmentation and feature extraction*

2.4.1. *Preprocessing*

Preprocessing of soil image means the enhancement of soil image quality for better feature extraction [38] and texture classification. Since the intention is to evaluate the soil classification using an Android Smartphone, soil images are collected using a Smartphone. The dimensions of the images are in 4160 × 2630 pixels with an average size of 2.3 MB. The contrasts of the images are enhanced using MatLab 2015a. The output of the contrast-enhanced image is presented in Fig. 16. Then the images are processed for noise removal and the low pass butter worth filter with a cut off frequency 0–0.5 is used to remove the noise of the images.

$$f = \frac{1}{1.0 + (w/\mathrm{cutoff})2n} \tag{iv}$$

Fig. 16 – Image Preprocessing and Segmentation.

Above Eq. (iv) shows the butter worth low pass filter for noise remove where n is the order of filtering. The order of the filtering is 2 in our approach. After the noise removal, the images are resized into a dimension of 300x400 and are processed for the extraction of the region of interest. The algorithm for image preprocessing is presented by Algorithm 1.

Algorithm 1: Algorithm for the Image Preprocessing
Input: Input Image **Output: Preprocessed Image** Step1: Read the Images Step 2: Enhanced the Contrast of the Image. Step 3: Remove the noise of the Contrasted Image using Low pass Butter worth Filter. Step 4: Images are resized in a dimension of 300x400.

2.4.2. *Image segmentaTion*

All the filtered images are with white background region. Though the images are with white background, the region of interest is extracted from the soil image using segmentation. Image segmentation is the process of segment an image into the different region and selects the region of interest of an image from it. It is found that the Gabor filter is a good model for texture segmentation of an image to identify the region of interest. For this approach, an array of Gabor filter is used with a different scale and orientations [12]. The scale and orientation of the Gabor filter for this approach are 4 and 6 respectively. For the implementation, a Matlab (R2015a) code is developed and the output of the code is figured out in Fig. 16. The first part of Fig. 16 represents the captured image, the second part of the Fig. 16 represents the contrast-enhanced image and the last part of the Fig. 16 represents the segmented image.

2.4.3. *Feature extraction of images*

Feature extraction is the process of extraction of different texture features of a soiled image. The extracted features of the image are presented as a feature vector of the soil images. The following steps are followed for the feature extraction and the method is presented in Fig. 17.

(a) In the first step, the Hue, Saturation, and Value (HSV) histogram [31] of the segmented images are calculated. The input image is quantized into Hue, Saturation, and Value (HSV) color space with $8 \times 2 \times 2$ equal bins. The output of histogram is a vector of size 1×32 these contains the Hue, Saturation, and Value of the images.
(b) In the second step, the color auto correlogram is calculated. It is used to find the spatial correlation of the identical pixels [23,31,32]. To implement color auto correlogram using MatLab 2015a, the input image is quantized into 64 colors in RGB color space as $4 \times 4 \times 4$ color space. The output is presented as a feature vector of size 1×64 containing color auto correlogram. This method follows a dynamic programming approach with a distance set $d = \{1\ 3\ 5\ 7\}$ as suggested by Hung et al. [23].
(c) In this step, the color moment of the input image is determined. It is used to find the color similarity between the images [24,31,32]. In this paper, the RGB color channel of the input image is considered and computed the 9 color Moments of the original image out of these 3 are for mean, 3 are for standard deviation [31] and 3 are for skewness. These moments are for the

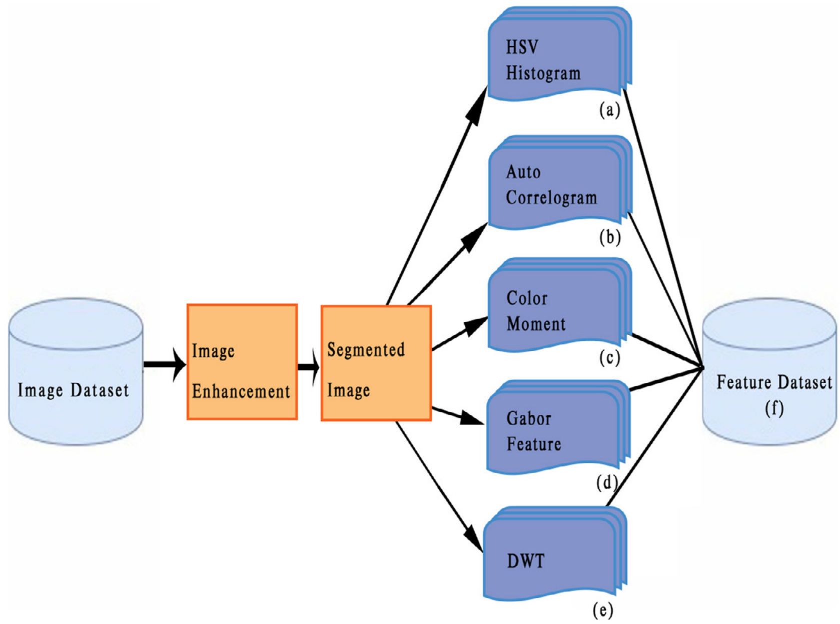

Fig. 17 – Flow Chart of Feature Extraction.

three channels of the images [13,29,32]. The output of this method is represented as a feature vector of size 1x9.

(d) In the fourth step, the Gabor feature of the soil images is calculated. Gabor wavelet is a method to find the Gabor feature of an input image [12,31,32]. This method is used to calculate the mean-squared energy and mean amplitude of the input image. For this paper, the number of wavelet scale is 4 and the number of filter orientation is 6. Before applying this method, images are converted into a grayscale format and perform the convolution using the First Fourier transformation. The output size of the mean-squared and mean-amplitude feature of the Gabor filter is 1×24 and 1×24 respectively. The final size of the output feature of the Gabor filter is 1×48.

(e) In the next step, the single level discrete 2-D wavelet decomposition [32] is performed. The wavelets coefficients of the input image are extracted. The feature vector contains the mean and standard coefficients of the discrete wavelet transform. The output size of the mean and standard coefficient feature of the discrete wavelet transform is 1×20 and 1×20 respectively. The final size of the output feature of the Discrete Wavelet Transform is 1×40.

(f) In the last step, all the calculated features are combined together to form the final feature vector of the image. The output of the combination is presented as the final feature vector of the methodology.

Initially, the multi SVM classifier is used to classify 3 different soil classes and later on, the multi SVM classifier is used to classify 12 different soil classes. The feature vector size of one single image is 1×193. The final feature vector size for 15 training images of 3 class soil classifier is 15×193. The size of the testing set is 13×193 for the entire 13 images of 3 class classifier. The final feature vector size for the entire 29 training images of 12 class classifier is 29×193 and the size of the testing set is 24×193.

Table 3 shows the numbers of soil sample of each class with the feature vector length of the image. Table 4 shows the mean feature vector value of the 50 soil samples.

The graphical representation of the feature vector of the different soil image is presented in Fig. 18.

3. Classification, result discussion, and analysis

Table 1 shows the percentage of sand, silt, and clay of different soil samples and results are displayed in terms of texture using USDA triangle method. It means that the hydrometer test gives only the percentage of sand, silt, and clay in the laboratory, but the final classification is done using USDA triangle. Table 5 shows the different features of the different soil images. The proposed system presented the feature vector in terms of a numeric value. This value is found from the texture properties of the soil images, so a correlation is found between the feature vector of the image and the texture of the soil. For that correlation, the final classifications of the soil samples are done using multi support vector machine and kernel method.

A support vector machine is a supervised binary classification technique [21,27,28]. It is proposed by Vapnik VN which is based on statistical learning theory [16,30]. The SVM is a very popular tool in machine learning and always gives a better result in texture classification [16,17,26,27], image retrieval and other classification problems [20,37]. Initially, it is used for binary classification where classes are separated with the help of a hyperplane. The pixel values nearest to this hyperplane are known as support pixel vector [19,22,27]. Later

Table 3 – Number of Soil Samples per class with feature vector length.

Class	Total Number of Sample	Feature Vector length one single image	Feature Vector length per class (Training)	Feature Vector length per class (Testing)
Clay Loam	7	1 × 193	4 × 193	3 × 193
Sandy Clay Loam	10	1 × 193	5 × 193	5 × 193
Silt Loam	11	1 × 193	6 × 193	5 × 193
Clay	4	1 × 193	2 × 193	2 × 193
Fine Sand	2	1 × 193	1 × 193	1 × 193
Loam	2	1 × 193	1 × 193	1 × 193
Loam Fine sand	1	1 × 193	1 × 193	1 × 193
Loamy Sand	1	1 × 193	1 × 193	1 × 193
Sandy Clay	3	1 × 193	2 × 193	1 × 193
Sandy Loam	5	1 × 193	3 × 193	2 × 193
Silt Clay	1	1 × 193	1 × 193	1 × 193
Silt Clay Loam	3	1 × 193	2 × 193	1 × 193
Total	**50**		**29 × 193**	**24 × 193**

Bold defines the Total Class, Total numbers of soil samples and total size of Training Set.

Table 4 – Mean value of Feature Vector for 50 Soil Samples.

Sample	Feature Vector	Sample	Feature Vector	Sample	Feature Vector
1	18.0094	18	11.3535	35	12.545
2	19.4434	19	20.1533	36	17.5665
3	18.5223	20	15.2958	37	17.9261
4	17.3816	21	16.8291	38	18.0012
5	19.012	22	15.9821	39	17.2351
6	12.222	23	23.4383	40	17.125
7	13.9614	24	17.982	41	16.145
8	13.912	25	14.9124	42	16.782
9	12.587	26	10.6254	43	13.7337
10	12.2535	27	12.0012	44	12.2106
11	16.9451	28	13.021	45	11.192
12	11.598	29	12.1201	46	19.324
13	10.291	30	12.225	47	19.0125
14	17.8785	31	12.0021	48	18.9871
15	10.856	32	12.1205	49	18.7871
16	10.9614	33	11.9851	50	19.821
17	12.1235	34	11.7584		

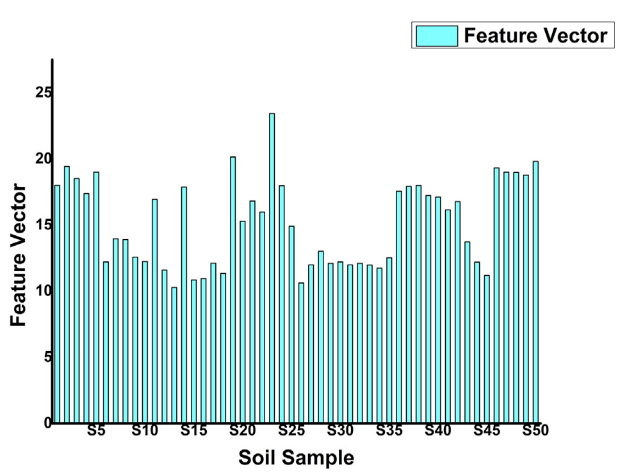

Fig. 18 – Feature Vector of Different Soil Sample.

on, it is extended for multi-class classification. In this paper, the multi SVM technique is used to classify the images with the help of linear kernel function as defined by mercer's theorem images [14,15]. Authors often used different machine learning techniques to classify the texture of the image using machine learning approaches such as Support Vector Machine, Artificial Neural network, Decision Tree, and KNN classification [19,22,27,28,34] but Support Vector machine often gives better result [20,37]. The accuracy of the different machine learning approach for the soil classification is presented in Table 7.

3.1. *Three class soil classification using multiSVM*

For this research, 50 soil samples are collected and the textures are classified using the hydrometer and USDA classification in the laboratory. The laboratory classification method

Please cite this article as: U. Barman and R. D. Choudhury, Soil texture classification using multi class support vector machine, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2019.08.001

Table 5 – Result of Three Class Classification (SVM).

Type	Average Percentage of accuracy
Clay Loam	97.70%
Sandy Clay Loam	96.21%
Silt Loam	93.25%
Total	**95.72%**

Bold defines the total average accuracy of 3 class classification.

gives 12 different classes of soils. It is found that more numbers of soil sample belong to Clay loam, Sandy Clay Loam, and Silt Loam. Total 7 soil samples belong to Clay, 10 soil samples belong to Sandy Clay Loam, and 11 soil samples belong to Silt Loam. In 2006, Bhattacharya B and Solomatine [8] used three classes Support Vector Machine for the classification of three soil classes and achieved 90.7% accuracy for the classification. In this paper, initially, Support Vector Machine classifier is used to classify three soil classes such as Clay loam, Sandy Clay Loam, and Silt Loam as these contain more number of soil samples. A total of 15 images are used for training and 13 images are used for testing. In this approach, the accuracy of the kernel based Multi Support Vector Machine is evaluated using cross-validation and the algorithm of the accuracy is presented in algorithm 4.

Multi Class Support Vector Machine classifies the class of the given training vector according to the given group and gives the result to which the class it belongs. Group in SVM is a vector which can be numeric, categorical and also sometimes it is logical. Each element in the group specifies the corresponding row of the training. In this three-class classification, the size of the training is 15x193 and group size is 1x15. The multi support vector machine classification is implemented using Matlab (R2015a). The training data and group data are used to create the support vector machine classifier structure (SVMStruct). Again support vector machine classifier structure (SVMStruct) and testing data are used to define the class to which the testing sample is assigned. The group vector for three class classification contains the values from 0 to 2 as the dataset contain only 3 soil classes. Now, based on the testing values, the classifier classifies the class of the test image as it is assigned. The pseudo code of the classifier result is presented in Algorithm 2.

Algorithm 2: Algorithm for the result of soil classification (3 Classes)

Input: multiSVM Result (test_result)
Output: Classification Result
Step 1: If (test_result == 0) then Class = Clay Loam.
Step 2: If (test_result == 1) then Class = Sandy Clay Loam
Step 3: If (test_result == 2) then Class = Silt Loam

3.2. *Twelve class soil classification using multiSVM*

In this step, Support Vector Machine class classifier is tested for entire 12 classes of soil. The size of the training is 29×193 and the group size is 1×29. The group vector contains the values from 0 to 11 as the soil dataset contains 12 different soil classes. Now, based on the testing values the classifier classifies the class of the test image as it is assigned. The pseudo code of the classifier result is presented in Algorithm 3.

Algorithm 3: Algorithm for the result of soil classification (12 Classes)

Input: multisvm Result (test_result)
Output: Classification Result
Step1: If (test_result == 0) then Class = Clay Loam.
Step 2: If (test_result == 1) then Class = Sandy Clay Loam
Step 3: If (test_result == 2) then Class = Silt Loam
Step 4: If (test_result == 3) then Class = Clay
Step5: If (test_result == 4) then Class = Fine Sand.
Step 6: If (test_result == 5) then Class = Loam
Step 7: If (test_result == 6) then Class = Loam Fine Sand
Step 8: If (test_result == 7) then Class = Loamy Sand
Step 9: If (test_result == 8) then Class = Sandy Clay
Step 10: If (test_result == 9) then Class = Sandy Loam
Step 11: If (test_result == 10) then Class = Silt Clay
Step 12: If (test_result == 11) then Class = Silt Clay Loam

3.3. *Accuracy for multiSVM*

The soils are classified using SVM with a kernel function. A kernel is a similarity function and it takes two inputs and spits out how similar they are [19]. In this paper SVM with a linear kernel is applied to classify the soil classes. The reason of using linear kernel is presented below.

a) The linear kernel is a very simple and faster kernel approach of SVM. The linear kernel is good when the images have too many features. This is because mapping the data to a higher dimensional space does not really improve the performance of the classifier [40]. In this paper, 5 different features of the soil images are extracted and it gives 193 different values of feature for a single image. It is found that the feature size of a single image is large and it is 1x193. So, the linear kernel with SVM is applied to classify the soil classes.
b) For this paper, non liner mappings of SVM are not performed for the dataset as it contains many data points for the feature vector.
c) In this paper, data are trained using SVM with linear kernel. Linear kernel is faster and simple kernel method. It requires optimizing the value of C parameter but others kernels require to optimize the value C and γ parameter. C in SVM is a regularization parameter which defines the cost of misclassification. It defines the bias and variance of the model. The high values of C mean low bias and high variance. The low values of C mean high bias and low variance. Low bias and high variance always lead to over-fitting of the model whereas high bias and low variance always lead to under-fitting of the model. In this paper, the value of C is considered as 1.0 and for this the model is neither over-fitted nor under-fitted.

The accuracy of the SVM classifier with the kernel is evaluated using the holdout method of cross-validation. During the time of accuracy calculation, holdout cross-validation of group vector is performed and creates a classifier performance object, these accumulates the result of the classifier for the testing vector. The algorithm for evaluating accuracy is presented by algorithm 4. So for evaluating the accuracy of the system, the kernel function is trained for 500 iterations and 1000 iterations. But the maximum accuracy of the system is found for 500 iterations. To overcome the model over fitting, the samples are divided into training and testing set as explained in Table 3 and the model is evaluated using cross-validation by setting the C parameter of kernel as 1.0. Cross-validation is re-sampling technique used to evaluate the SVM model as less numbers of samples are present in the soil dataset. The average testing accuracy of the mode is more than 90% which is defined in Tables 5 and 6.

Algorithm 4: Algorithm for Accuracy of Classification

Input: Training Data, Group Data
Output: Accuracy in percentage
Step1: Consider iteration = 500
Step 2: For i = 1 to iteration
Step 2.2: Take Train and load Group vector
Step 2.3: Perform Holdout Cross-Validation of Group Vector.
Step 2.4: Create Classifier performance object, CP for Group vector.
Step 2.5: Perform SVM training to create the SVM structure using Linear Kernel Function, Training Vector and Group Vector.
Step 2.6: Now, Calculate the test_result by using Group Vector and Test Vector.
Step 2.7: Again Create Classifier performance object, CP for Test vector by considering test_result and calculate the accuracy.
Step 2.8: Print the accuracy of the kernel in percentage.

For the three class classification, multi Support Vector Machine classifier gives an average of 95.72% of accuracy. The results obtained from the multi SVM classifier for the three classes are presented in Table 5. Multi SVM gives good accuracy for the three class classifications because the soil classes such as Clay Loam, Sandy Clay Loam, and Silt Loam have more soil samples as compared to the other soil class. Different soil samples are used for training and testing and it is presented in Table 3. In the next step, Multi SVM is tested for the entire 12 soil classes and the accuracy is presented in Table 6.

From Tables 5 and 6, it is found that the average accuracy of 12 class SVM classifier is less as compared to the 3 class SVM classifier. It is because of the less numbers of soil sample in the Loam Fine sand, Loamy sand, and Silt Clay soil. In 3 class classifier, the number of samples of Clay Loam, Sandy Clay Loam, and Silt Loam are 7, 10, and 11 respectively. So, the variations of soil samples are not much more to each others. But in 12 class classifier, only 1 sample is present in Loam Fine sand, Loamy Sand, and Silt Clay respectively. So, the variations of these three classes are more as compared to the Clay Loam, Sandy Clay Loam, and Sandy Loam, Clay, Silt Clay Loam and Silt Loam soil. It leads to imbalanced dataset for these three classes. But the dataset is not imbalanced with respect to the Fine Sand, Loam soil as these two classes have only 2 soil samples. Though the samples of Loam Fine sand, Loamy sand, and Silt Clay classes are less in number, the accuracy of the model is evaluated using the cross-validation by setting the C parameter of kernel based SVM as 1.0 and achieved more than 80% accuracy for the classes. As the accuracy of these three classes is less as compared to others, the overall accuracy of the model is less from classifying 3 to 12 classes.

Table 6 – Result of Multi-Class Classification (SVM).

Type	Average Percentage of accuracy
Clay Loam	96.20%
Sandy Clay Loam	92.56%
Silt Loam	95.25%
Clay	90.54%
Fine Sand	92.58%
Loam	96.84%
Loam Fine sand	85.54%
Loamy Sand	81.25%
Sandy Clay	92.56%
Sandy Loam	95.65%
Silt Clay	84.25%
Silt Clay Loam	93.21%
Total	**91.37%**

Bold defines the total average accuracy of the 12 class classifier.

3.4. *Result analysis of multiSVM*

Initially, three classes of soils are classified using three-class SVM classifier with linear kernel and the accuracy of the classifier is evaluated using cross-validation. The accuracy of the classifier is best for Clay loam soil and is 97.7%. But the average accuracy of the classifier is 95.72%. Bhattacharya B and Solomatine [8] found an average of 90.7% accuracy for SVM class classifier in case of sand, clay and peat type of soil. Wu W et al [27] found the 79.4% accuracy for Clay soil, 99.2% accuracy for Loam soil, 66.1% accuracy for Sand soil.

The SVM classifier gives a good result for three classes of soil, so it is applied for all 12 classes of soil. The results of the multi SVM classification for the 12 classes are presented in Table 6. The classifier gives more than 90% accuracy for all types of soil class except the loamy fine sand, loamy sand, and silty clay. Srunitha and Padmavathi [21] applied seven classes SVM classifier for seven different types of soil samples and found 60.9% accuracy for correctly classified soil segments and 74.7% for correctly classified instances. Vibhute et al. [22] found an average accuracy of 71.18% for five types of soil samples. They did not find the efficient accuracy with SVM as due to the complex nature of the selected region and spectral information of the soil samples. Guang Y et al. [30] achieved 96.67% accuracy for the seven class classification using multi SVM with a linear kernel function. In this paper, the accuracy of the multi-class SVM carries out with the help of linear kernel function with 500 iterations. The multi SVM class classifier gives the best testing accuracy result for the Loam and Clay Loam soil and it is respectively 96.84% and 96.2%. The average testing accuracy of the entire 12 class classifier is 91.37%. The accuracy for Loam Fine sand, Loamy sand, and Silt Clay are respectively 85.54%, 81.25%, and

Please cite this article as: U. Barman and R. D. Choudhury, Soil texture classification using multi class support vector machine, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2019.08.001

Table 7 – Comparative Analysis of Previous Research Methods and Proposed Method.

Model	Number of Soil Class for the Experiment	Classification Algorithm	Accuracy
Bhattacharya and Solomatine [8]	3	Multi SVM with Liner kernel	90.7%
Chung SO et al. [9]	13	Linear Regression	48%
Srunitha and Padmavathi [21]	7	Multi SVM with Liner kernel	60.9%
Vibhute et al. [22]	5	Multi SVM with Liner kernel	71.78%
Wu W et al [27]	3	Multi SVM with Polynomial Kernel	0.794 for clay, 0.992 for loam, 0.661 for sand.
Zhao Z et al. [26]	2	ANN	88% for Clay 81%. for Sand
Guang Y et al. [30]	7	Multi SVM with Polynomial Kernel	96.67%
Guang Y et al. [30]	7	Partial Least Squares Discriminant Analysis (PLS-DA)	93.33%
Mengistu AD and Alemayehu DM [33]	6	Back-Propagation Neural Network (BPNN)	89.7%
Chandan, Thakur R [35]	7	Linear SVM	99%
Chandan, Thakur R [35]	7	Fine KNN	94.8%
Chandan, Thakur R [35]	7	ANN	94.2%
Proposed Model	3	Multi SVM with Liner kernel	95.72%
Proposed Model	12	Multi SVM with Liner kernel	91.37%

84.25% and it is less as compare other soil classes because the same image used as testing and training. The accuracy of the different machine learning method is compared with the accuracy of the proposed approach and is presented in Table 7.

The proposed system is tested by collecting another 5 soil samples from the west zone of the Guwahati with the help of rural farmers and then soil images are captured using the same android mobile phone in front of the farmers. These 5 samples are labeled as Sample 1, Sample 2, Sample 3, Sample 4 and Sample 5. Out these 5 samples, sample no. 1 is Clay loam, Sample no. 2 is Loam, Sample no. 3 and Sample no. 4 is Silt Loam and Sample no. 5 is Clay. The soil samples are tested in the laboratory using the hydrometer method and their textures are determined using USDA triangle. Then, the features of the soil samples are determined using the proposed method and textured are classified with SVM. The proposed system gives the same texture as the hydrometer method for those 5 samples and their accuracy is more than 90% for each soil sample. Table 8 shows the comparative analysis of the hydrometer method and the proposed method for that 5 soil class.

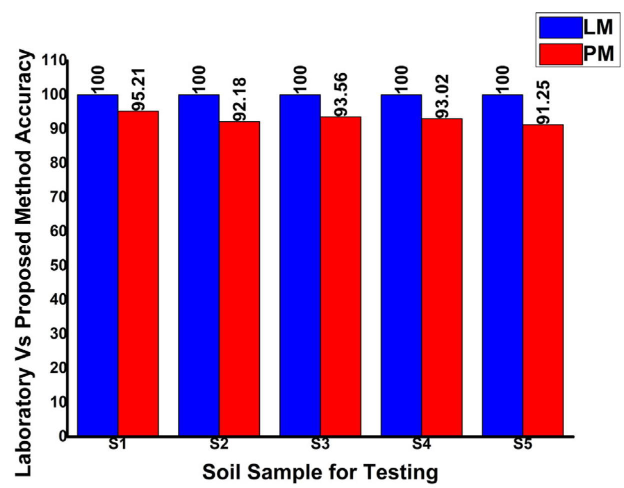

Fig. 19 – Comparison of Result.

Fig. 19 shows the graphical comparative analysis of the laboratory method (LM) and proposed method (PM) for 5 types of selected soil samples.

Table 8 – Comparative Analysis of Laboratory Method and Proposed Method.

Laboratory System		Proposed Method		Average Percentage of accuracy
Sample No	Texture	Sample	Texture	
1	Clay Loam	1	Clay Loam	95.21%
2	Loam	2	Loam	92.18%
3	Silt loam	3	Silt loam	93.56%
4	Silt loam	4	Silt loam	93.02%
5	Clay	5	Clay	91.25%

Please cite this article as: U. Barman and R. D. Choudhury, Soil texture classification using multi class support vector machine, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2019.08.001

4. Conclusion

In this paper, an image based texture analysis is presented to classify the soil images using multi SVM and linear kernel function. The images are captured using android mobile phone camera within West Guwahati Region. The three class classifier and multi-class classifier gives a good performance for the real dataset except for the loamy fine sand, loamy sand, and silty clay. Earlier, the texture of the soil is determined with the traditional hydrometer method and USDA triangle which is a very time and labour consuming process. A simple test of hydrometer takes at least 24 h for the percentage calculation of sand, silt, and clay. But the proposed system is accurate and takes less time to classify the soil. It gives a fast and accurate result for soil classification with the help of a support vector machine an android Smartphone.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgment

For this work, we are giving our heartiest thanks to the Department of Civil and Computer Science & Engineering Girijananda Chowdhury Institute of Management and Technology, Guwahati.

REFERENCES

[1] Sun Y, Long Z, Jang PR, Plodinec MJ. Gabor wavelet image analysis for soil texture classification. In: Proc. SPIE 5587, Nondestructive Sensing for Food Safety, Quality, and Natural Resources. p. 254–61.

[2] Bae SG, Yeon IK, Park SD, Kang CK, Zakaullah K. Effects of soil texture by soil addition on the growth and quality of oriental melon (Cucumis melo L. var. makuw Mak.) under protected cultivation. J Bio-Environ Con 2004;13(3):156–61.

[3] Breul P, Gourves R. Field Soil Characterization: Approach Based on Texture Image Analysis. J Geotech Geoenviron Eng 2006;132(1).

[4] Zhang X, Younan NH, King RL. Soil texture classification using wavelet transform and maximum likelihood approach. In: IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Toulouse: Proceedings (IEEE Cat.No.03CH37477); 2003. p. 2888–90.

[5] Zhang X, Vijayaraj V, Younan NH. Hyperspectral soil texture classification. In: IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003, Greenbelt, MD, USA. p. 182–6.

[6] Zhang X, Younan NH, O'Hara CG. Wavelet domain statistical hyperspectral soil texture classification. IEEE Trans Geosci Remote Sens 2005;43(3):615–8.

[7] Shenbagavalli R, Ramar K. Classification of soil textures based on Law's features extracted from preprocessing images on sequential and random windows. Bonfring Int J Adv Image Process 2011;1(1):15–8.

[8] Bhattacharya B, Solomatine DP. Machine learning in soil classification. Neural Network. 2006;19(2):186–95.

[9] Chung SO, Cho KH, Cho JW, Jung KY, Yamakawa T. Soil texture classification algorithm using rgb characteristics of soil images. J. Fac. Agr. kyushu univ 2012;57(2):393–7.

[10] Gee GW and D Or. Particle size analysis. In: Dane JH and Top GC, editors. Methods of soil analysis: SSSA, Madison, Wisconson; 2002. p. 255–293.

[11] Bouyoucos GJ. Hydrometer method improved for making particle size analysis of soils. Agron J 1962;54:464–5.

[12] Jain AK, Farrokhnia F. Unsupervised Texture Segmentation Using Gabor Filters. J Pattern Recognition 1991;24(12):1167–86.

[13] Ahmed JA, Wesam MA. Image retrieval based on content using color feature. ISRN Comput Graphics 2012:11.

[14] Cristianini N, Shawe-Taylor J. Support Vector Machines. In An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge: Cambridge University Press; 2000. p. 93–124.

[15] Kecman V, Huang TM, Vogt M. Iterative single data algorithm for training kernel machines from huge data sets: Theory and performance. In: Wang L, editor. Support vector machines: Theory and applications. Studies in Fuzziness and Soft Computing, vol. 177. Berlin, Heidelberg: Springer; 2005. p. 255–274.

[16] Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer; 1995.

[17] Pai PF. System reliability forecasting by support vector machines with genetic algorithms. Mathemat Comput Model 2006;43(3–4):262–74.

[18] Soil Texture Calculator, NRCS Soils USDA. Link: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_054167.

[19] Tian D, Zhao X, Shi Z. Support vector machine with mixture of kernels for image classification. In: Proc. of the 7th Int'l Conf. on Intelligent Information Processing (ICIIP'12). p. 67–75.

[20] Wang X, Yang H, Li Y. A new SVM-based active feedback scheme for image retrieval. Eng Appl Artif Intell 2015;37:43–53.

[21] Srunitha K, Padmavathi S. Performance of SVM classifier for image based soil classification. In: IEEE Conference on Signal Processing, Communication, Power and Embedded System (SCOPES). Paralakhemundi. p. 411–5.

[22] Vibhute AD, Kale KV, Dhumal RK, Mehrotra SC. Soil type classification and mapping using hyperspectral remote sensing data. In: Conference on Man and Machine Interfacing (MAMI). Bhubaneswar. p. 1–4.

[23] Haung J, Kumar SR, Mitra M, Zhu J, Zabih R. Image indexing using color Correlograms. In: Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition. San Juan, Puerto Rico, USA. p. 762–8.

[24] Yu H, Li M, Zhang H, Feng J. Color texture moment for content based image retrieval. In: . Proc. International Conference on image processing. Rochester, NY, USA. p. 929–32.

[25] Guwahati. Link: https://goo.gl/maps/GDZ6mQFnqJt.

[26] Zhao Z, Chow TL, Rees HW, Yang Q, Xing Z, Meng FR. Predict soil texture distributions using an artificial neural network model. Comput. Electron. Agric. 2009;65:36–48.

[27] Wu W, Li AD, He XH, Ma R, Liu HB, Lv JK. A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. Comput Electron Agric 2018;144:86–93.

[28] Kovačević M, Bajat B, Gajić B. Soil type classification and estimation of soil properties using support vector machines. Geoderma 2010;154:340–7.

[29] Patel AK, Chatterjee S. Computer vision-based limestone rock-type classification using probabilistic neural network. Geosci Front 2016;7:53–60.

[30] Guang Y, Shujun Q, Pengfei C. Rock and soil classification using PLS-DA and SVM combined with a laser-induced breakdown spectroscopy library. Plasma Sci Technol 2015;17(8):656–63.

Please cite this article as: U. Barman and R. D. Choudhury, Soil texture classification using multi class support vector machine, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2019.08.001

[31] Mistry Y, Ingole DT, Ingole MD. Content based image retrieval using hybrid features and various distanc metric. J Electr Syst Inform Technol 2017;5(3):874–88.

[32] Vinayak V, Jindal S. CBIR system using color moment and color auto-correlogram with block truncation coding. Int J Comput Appl 2017;161(9):1–7.

[33] Mengistu AD, Alemayehu DM. Soil characterization and classification: a hybrid approach of computer vision and sensor network. Int J Electrical Comput Eng 2018;8(2):989–95.

[34] Huang T, Yang R, Hunag W, Qaio X. Detecting of sugarcane borer disesaes using support vector machine. Informat. Proc. Agricult. 2018;5:74–82.

[35] Chandan T, Thakur R. An intelligent model for indian soil classification using various machine learning techniques. Int J Comput Eng Res 2018;8(9):33–41.

[36] Chung SO, Cho KH, Kong JW, Sudduth KA, Jung KY. Soil texture classification algorithm using RGB characteristics of soil images. IFAC Proc Vol 2010;43(26):34–8.

[37] Kanimozhi T, Latha K. An integrated approach to region-based image retrieval using firefly algorithm and support vector machine. Neurocomputing. 2015;151(3):1099–111.

[38] Singh V, Misra A. Detection of plant leaf diseases using image segmentation and soft computing techniques. Inf Process Agric 2017;4(1):41–9.

[39] Barman U, Choudhury RD, Talukdar N, Deka P, Kalita I, Rahman N. Predication of soil pH using HSI colour image processing and regression over Guwahati, Assam, India. J Appl Natural Sci 2018;10(2):805–9.

[40] SVM: https://www.svm-tutorial.com/2014/10/svm-linear-kernel-good-text-classification/.

Please cite this article as: U. Barman and R. D. Choudhury, Soil texture classification using multi class support vector machine, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2019.08.001