

# An Intelligent System for Soil Classification using Unsupervised Learning Approach

Olanloye, Dauda Odunayo

Emmanuel Alayande College of Education, Oyo. PMB 1010, Oyo, Oyo State, Nigeria

## ABSTRACT

The traditional soil analysis technique when applied is time consuming, labour intensive and expensive. The research made an attempt to develop an intelligent system that is capable of classifying soil in a particular location if the hyperspectral data of such location is available. The system was developed using unsupervised learning. Wavelet transform was used to denoise the spectral signal at preprocessing stage. Fuzzy c- means was used for clustering in order to identify the cluster centre. KSOM is applied for the purpose of classifying soil into various classes. The system was implemented using R programming language.

## Keywords

Intelligent System, Hyperspectral Data, Spectral, Fuzzy C-means, KSOM, Cluster Centre, Wavelet Transform

## 1. INTRODUCTION

There are different ways of defining an intelligent system. For instance, someone that has the ability to learn very fast or the ability to remember the past event could be called an intelligent person. However, for the purpose of this work, intelligent is the ability to learn and apply the knowledge to solve some related problem.

A system is the combination of different components which works together to achieve a particular task.

Artificial intelligence may be defined as the branch of science that is concerned with automation of intelligent behaviour [20]. This definition suffers a lot of setback as it is very difficult to define artificial intelligence without the full understanding of the word intelligence. It is therefore doubtful that anyone can come to defining intelligence in a way that will be specified enough to help in evaluation of a supposedly intelligent computer program while still capturing the validity of the human mind.

Thus, the problem of defining artificial intelligence became one of defining intelligent itself. What is intelligence? Intelligence could be defined as ability to learn, understand and being able to apply the knowledge acquired to solve similar problem in future.

Psychologist and cognitive theorist are of the opinion that intelligence helps in identifying the right piece of knowledge at the appropriate instance of decision making [2].

Amit [2] defined Artificial Intelligence as simulation of human intelligence on a machine so as to make the machine efficient to identify and use the right piece of “knowledge” to solve a particular problem.

Intelligence is the capacity of a system to achieving a goal or sustains desired behaviour under condition of uncertainty. Intelligent systems have to cope with sources of uncertainty like occurrence of unexpected event such as unpredictable changes in the world in which the system operate and incomplete, inconsistent and unreliable information available

to the system which operates for the purpose of deciding what to do next.

Intelligent system exhibits intelligent behaviour. Intelligent behaviour if exhibited is capable of achieving specified goals or sustains desired behaviour under conditions of uncertainty even in a poor structured environment.

According to Nikola [14], an intelligent system exhibits the following behaviour.

- (i) They should from time to time accommodate a new problem solving rule.
- (ii) They should be able to analyze themselves in term of behaviour, error and success.
- (iii) Once they are to interact, they should learn and improve through interaction with the environment.
- (iv) They should learn quickly from large amount of data.
- (v) They should have many base exemplar storage and retrieval capability.
- (vi) They should have parameter to present.

Agris [1] summarized basic features of intelligent system as follows:

- (i) They have ability to generate a new knowledge from already existing ones
- (ii) They have ability to learn
- (iii) They have ability to sense environment
- (iv) They have ability to act.

Different types of intelligent system have been developed and series of real life problems have been solved using intelligent systems. But, enough has not been done in the area of Agriculture most especially in the area of developing an intelligent system for soil classification.

This research work made an attempt to develop an intelligent system for soil classification using unsupervised learning approach.

Although, soil is considered as just thin layers of surficial unconsolidated material, they are a vital component of an interconnected ecosystem that influences every landscape [18].

The variability of soil properties across a landscape can influence habitat types which then shape the distribution of different animal species [18]. It has been suggested that as a fundamental land resources, soil productivity has influence the economy and development of many countries and hence “the advancement of the modern world” [18]. But when the soils are degraded, such as through poor Agricultural practices, it has been shown that the entire civilization can collapse [21].

It is therefore necessary to develop a tool that will be able to provide adequate information for maintaining the performance of our soil from time to time. The traditional soil analysis technique requires time intensive methods which becomes limiting when applied at regional and global scale [4]. Likewise, the earlier research works are not adequately providing enough information to monitor and maintain our soil.

This research work is to provide an alternative tool to inexpensively, rapidly and accurately classify different types of soil with an intelligent system using unsupervised learning approach. The intelligent system will be implemented with R programming language and make use of hyperspectral data.

## **2. LITERATURE REVIEW**

Stephan et al [18] carried out a research work on hyperspectral analysis of soil Nitrogen. The research was able to characterize the soil attribute using hyperspectral sensor with regression tree method.

Sergio [17] used Artificial Neural Network (ANN) for digital soil mapping. The research work made an attempt to compare Multilayer Perception (MLP) and Self Organizing Map.

Zhengyang [22] also predict soil texture distribution using ANN based on soil attribute obtained from existing coarse resolution soil maps combined with hydrographic parameter. The calibration could be used to produce soil maps in area with similar conditions without additional field survey.

In the past, there was a general review on remote sensing of soil properties. With the launch of Landsat in 1972, millions of multispectral images were made available which some researchers used to estimate soil properties [[6][8][15].

The method of using hyperspectral data to predict major soil properties has been established [5][6][10][19]. Literature reveals that despite series work earlier published, no attempt has been made to develop an intelligent system for soil classification using hyperspectral data and implemented with R programming.

## **3. STATEMENT OF THE PROBLEM**

Lack of information pertaining to soil distribution and maintenance contribute to uncertainties of predicting food production and lack of reliable and harmonized soil data has considerably limited environmental impact assessment, land degradation studies and adapted sustainable land management intervention [13].

There is also lack of consistency across countries concerning soil classification and legend. Unfortunately, various researchers have not been able to provide enough solution to this problem. The proposed intelligent system will be able to carry out soil classification in any part of the world if the hyperspectral data is made available [9].

## **4. SIGNIFICANT OF THE STUDY**

Soil is a gift of nature which supports all living things or creatures. Without soil there wouldn't be life. Soil is useful in different ways. For instance, in civil engineering, it is the gift of nature that holds the foundation of construction works. Therefore individuals, schools, institutions, industries, company buildings have their foundation properly gripped by different soil layer (rocks). Farmers see soil as a natural gift that host plants and crops. The farmland is cultivated on the soil. Again, it supplies food and nutrient to all crops or plants. The mining industries see soil as house / host of minerals. Soil

is needed by both plants and animals (including man) for sustenance of life

Despite various types of soil surveying being carried out in different countries, the scale and spatial covering of many conventional soil maps are not sufficient enough for planning and maintaining soil facilities at national and international level.

Therefore, for the benefit of mankind, there should be a tool that provides adequate information about different classes of soil to aid decision making in soil maintenance and management.

## **5. OBJECTIVES OF THE STUDY**

The objectives of the research are:

- (i) To develop an intelligent system for soil classification
- (ii) To develop the intelligent system using unsupervised learning approach
- (iii) To implement the intelligent system with hyperspectral data using R programming language.

## **6. METHODOLOGY**

In this research work, hyperspectral data was obtained, from which various reflectance bands were generated (fig 3). The spectrum of a pixel is thereafter obtained as shown in fig 4. A halving algorithm and fuzzy-c clustering were used to generate three (3) different cluster centres (fig 5).

Unsupervised learning approach was adopted using Kohonen Self Organizing Map with the 3 cluster centres as input during the learning process. The KSOM was able to classify the soil into different classes as shown in the soil predicted map depicted in fig 7. The training progress of the KSOM was shown in fig 8 while fig 9 and 10 shows the plot of learning SOM and class count plot of trained SOM. The system was implemented using R programming Language.

## **7. SUPERVISED AND UNSUPERVISED LEARNING**

### **7.1 Supervised Learning**

Supervised Learning has been applied to solve series of real life problems with great success. It can also be referred to as classification or inductive learning. Supervised learning is synonymous to human learning. Man learns from previous knowledge or experience to solve a particular problem. Likewise machines learn from the pattern of previous data which are collected in the past and represent past experience in some real life applications.

### **7.2 Unsupervised Learning**

Unsupervised learning is much more difficult compared to supervised learning. In this type of learning, the system is made to do some things but instructions on how to do it were not provided. There are two different approaches. The first approach is to use reward system to indicate success. This is not actually to produce classification but to make decisions that maximize reward. So it is applicable when agents are rewarded or compensated for carrying out certain actions and punished for doing others. This approach nicely generalizes to the real world, where agents might be rewarded for doing certain actions and punished for doing others.

A second type of unsupervised learning is called clustering. In this type of learning, the goal is not to maximize a utility function, but simply to find similarities in the training data. The assumption is often that the clusters discovered will match reasonably well with an intuitive classification [23].

### 7.2.1 KSOM

Self organizing maps are competitive neural network where the neurons are organized in two dimensional grids. Going by the learning rule, it has been established that vectors that are similar to each other in a multidimensional space will also be similar in two dimensional spaces. Self organizing maps are basically used for data classification.

#### 7.2.1.1 The KSOM Algorithm

**Step 1:** Set up input neuron matrix, In \_ X and In \_ Y.

Thus, total number of input neurons,

$$I = \text{In\_X} * \text{In\_Y}.$$

We used  $i = 0$  to  $I-1$  for numbering the neurons in this layer.

Eg.,  $X_j$  is the label for the input neurons i.e.  $X_0$  to  $X_{I-1}$

**Step 2:** Set up competitive layer matrix, Out \_ X and Out \_ Y.

For simplicity  $\text{Out\_X} = \text{Out\_Y}$  (such that we have a square map).

Therefore, total number of competitive neurons,

$$J = \text{Out\_X} * \text{Out\_Y}$$

We used  $j = 0$  to  $J-1$  for numbering the neurons in this layer

**Step 3:** Initialize connection weights (randomize) between input layer neurons and competitive layer neurons,  $W_{ij}$ .

Set initial topological neighborhood parameters,  $d_0$ .

Usually,

$$d_0 \leq \frac{\text{Out\_X}}{2}$$

Set initial learning rate parameter,  $\alpha_0$

(usually between 0.2 to 0.5)

Set total number of iterations,  $T$  (usually 10,000).

Start with iteration  $t = 0$ .

Apply the first pattern to the input of the KSOM.

**Step 4:** Compute the winning neuron ( $j$ ) in the competitive layer shown in fig 1 which is the minimum Euclidean distance from input layer to competitive layer such that

First, for each  $j$  (from  $j = 0$  to  $J-1$ ), compute the Euclidean distance as follows:

$$\|E(j)\| = \sqrt{\sum_{i=0}^{I-1} (w_{ij} - x_i)^2} \quad \dots\dots (1)$$

Then compare all these distances i.e. from

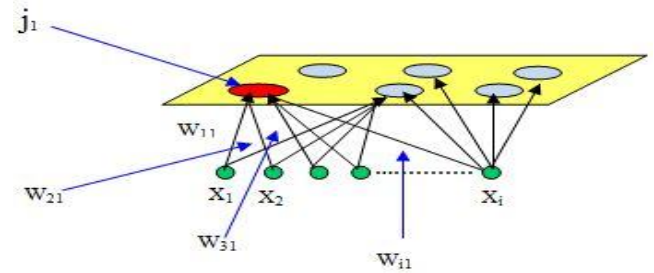
$$\|E(0)\| \text{ to } \|E(j)\|$$

and find the minimum distance  $\|E(j_c)\|$  which is the winner neuron,  $j_c$ .

$$\|E(j_c)\| = \min \|E(j)\|$$

Calculate Euclidean distance for each Competitive layer neuron

$$\|E(j_1)\| = \sqrt{(w_{11} - x_1)^2 + (w_{21} - x_2)^2 + \dots + (w_{i1} - x_i)^2} \dots (2)$$



**Fig 1: Diagrammatic illustration of the KSOM Algorithm**

**Step 5:** Update weight for each connections i.e. For all neurons  $j$  within a specified neighbourhood of  $J$ , and for all  $i$  :

$$W_{ij}(\text{new}) = w_{ij}(\text{old}) + \Delta w_{ij}(\text{new}) \quad \dots\dots (2)$$

Where

$$\Delta W_{ij} = \begin{cases} \alpha_t (x_i - w_{ij}(\text{old})) & \text{if unit is in neighbourhood } d_t \\ 0 & \text{otherwise} \end{cases}$$

**Step 6:** Update learning rate  $\alpha_t$  such that :

$$\alpha_t = \alpha_0 \left(1 - \frac{t}{T}\right) \quad \dots\dots (3)$$

**Step 7:** Reduce radius of topological neighborhood at specified times:

$$d_t = \text{int}[d_0 \left(1 - \frac{t}{T}\right)] \quad \dots\dots (4)$$

**Step 8:** Increase iteration  $t: t=t+1$

Repeat Steps 5 to 8 until  $t=T$

**Step 9:** Repeat with next pattern chosen randomly

(Do Steps 4-9)

#### 7.2.1.2 Computer Simulations of Self-Organization in the Kohonen SOM

1. Initialize weights to  $0.5 + 10\%$  randomized value.
2. 2 input vectors,  $X_1$  and  $X_2$  with several scores of entries between the range of 0 and 1.
3. Fig 1 given shows a plot of initial weights,  $w_{ij}$ .
4. Each unit in the competitive layer shows a point on the graph
4. Each unit in the competitive layer shows a point on this graph
5. The coordinate values of this point are the values of the incoming weights for the unit, thus  $w_{i1}$ ,  $w_{i2}$  are plotted for each competitive unit  $j$ .
6. All pairs of units in the competitive layer that are adjacent are connected.
7. This will allow us to see how the pattern of weights changes as the network organization evolves during training.

### 7.2.1.3 Some Applications of the KSOM

KSOM has been used for solving a variety of pattern classification problems.

Prof. Teuvo Kohonen has used the KSOM in developing a phonetic typewriter, a speech recognition device. Other applications include classification of images for remote sensing application, character recognition, etc.

### 7.2.1.4 Advantages of KSOM

- The KSOM (unsupervised learning ANN) is trained without teaching signals or target.
- Based on a series of input patterns, KSOM learns by itself to cluster the patterns according to their similar features.
- It can discover the existence of unlabelled data cluster.

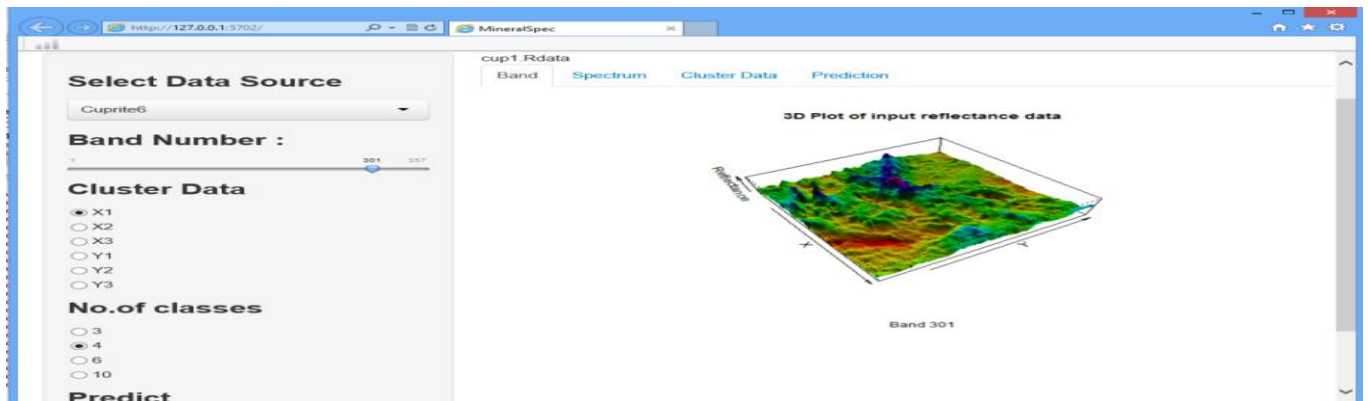


Fig 2: Main interface of the system

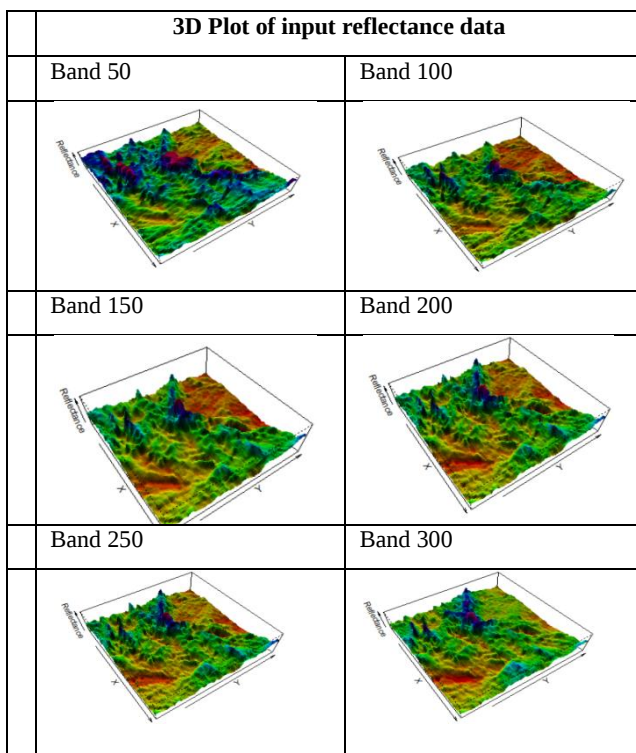


Fig 3: 3D Plot of input reflectance data

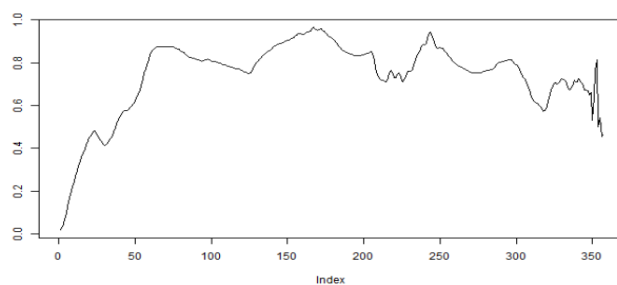


Fig 4: The spectrum of a pixel

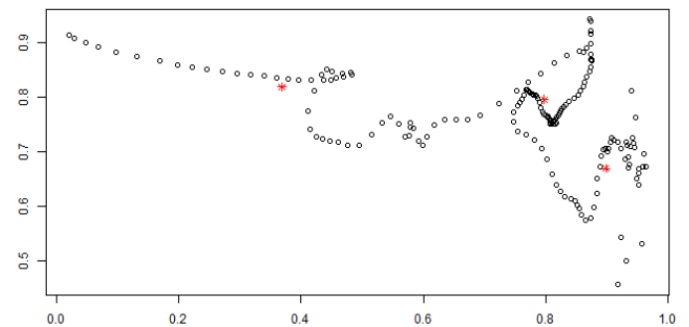


Fig 5: Clustered image of a pixel with 3 Cluster centers shown in red

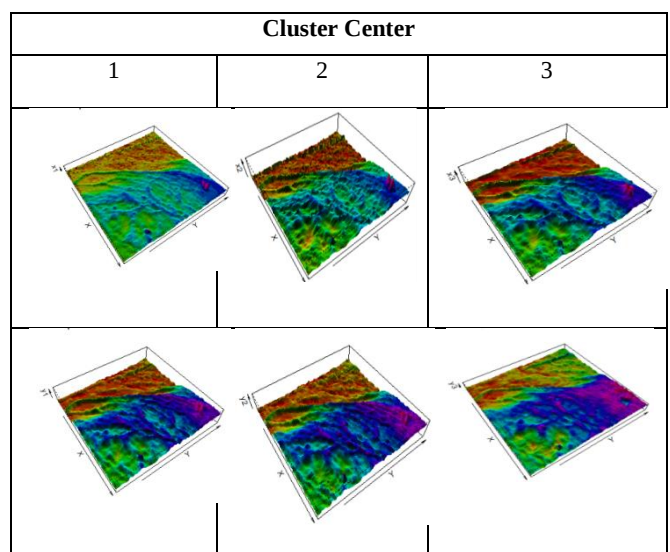


Fig 6: 3D Plots of the cluster center data

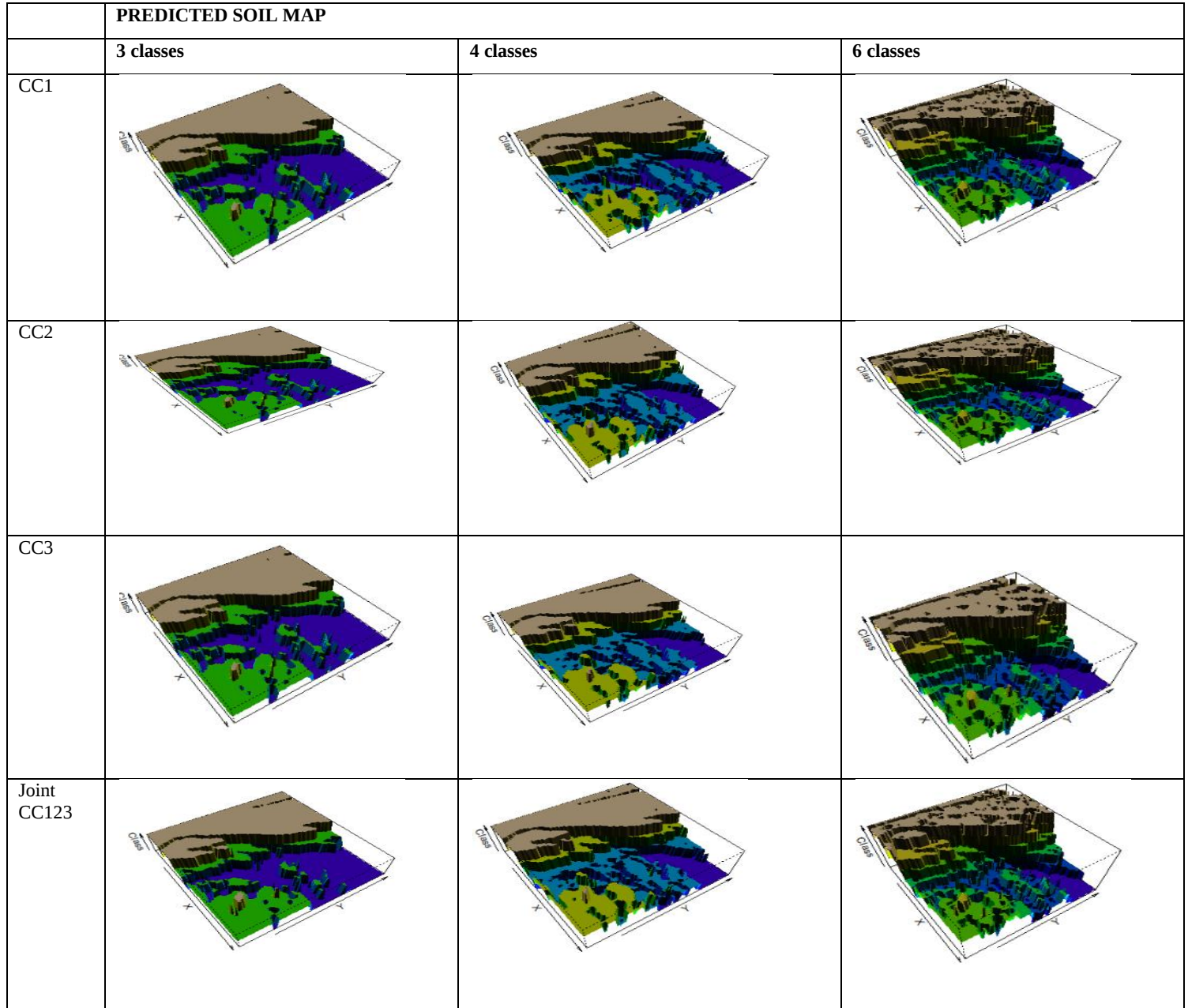


Fig 7: 3D Plots of the predicted soil maps

Table 1: Cluster center data in table form

| X1       | X2       | X3       | Y1       | Y2       | Y3       | Class |
|----------|----------|----------|----------|----------|----------|-------|
| 0.187544 | 0.564882 | 0.679369 | 0.641074 | 0.662144 | 0.725365 | 8     |
| 0.18795  | 0.567117 | 0.680288 | 0.641872 | 0.664031 | 0.728318 | 8     |
| 0.141356 | 0.474444 | 0.666162 | 0.62406  | 0.649036 | 0.711058 | 8     |
| 0.139944 | 0.474134 | 0.678972 | 0.627369 | 0.661414 | 0.724067 | 8     |
| 0.201181 | 0.609343 | 0.694761 | 0.642069 | 0.675314 | 0.756747 | 7     |
| 0.200752 | 0.607315 | 0.702623 | 0.642655 | 0.673217 | 0.757339 | 7     |
| 0.202342 | 0.602253 | 0.699838 | 0.649851 | 0.672485 | 0.754533 | 7     |
| 0.196241 | 0.587546 | 0.688304 | 0.644239 | 0.663774 | 0.743914 | 7     |
| 0.194845 | 0.580307 | 0.679101 | 0.633244 | 0.660266 | 0.739818 | 8     |

|          |          |          |          |          |          |   |
|----------|----------|----------|----------|----------|----------|---|
| 0.193765 | 0.587714 | 0.681924 | 0.622237 | 0.648373 | 0.731909 | 8 |
| 0.208384 | 0.630368 | 0.711756 | 0.60313  | 0.633919 | 0.735053 | 6 |
| 0.232483 | 0.666517 | 0.750961 | 0.624342 | 0.672428 | 0.801072 | 4 |
| 0.223267 | 0.646287 | 0.750033 | 0.643923 | 0.697677 | 0.824638 | 5 |
| 0.207427 | 0.629786 | 0.727853 | 0.635386 | 0.674351 | 0.78016  | 5 |
| 0.213699 | 0.647265 | 0.735844 | 0.624017 | 0.65978  | 0.770755 | 6 |
| 0.235871 | 0.676091 | 0.742273 | 0.631429 | 0.682011 | 0.811446 | 5 |
| 0.239554 | 0.667786 | 0.744385 | 0.6313   | 0.685634 | 0.811291 | 5 |
| 0.241672 | 0.664984 | 0.74817  | 0.63928  | 0.693744 | 0.813066 | 5 |
| 0.251334 | 0.684509 | 0.743694 | 0.642469 | 0.711104 | 0.833196 | 5 |
| 0.257009 | 0.695024 | 0.759544 | 0.657884 | 0.729256 | 0.85309  | 4 |

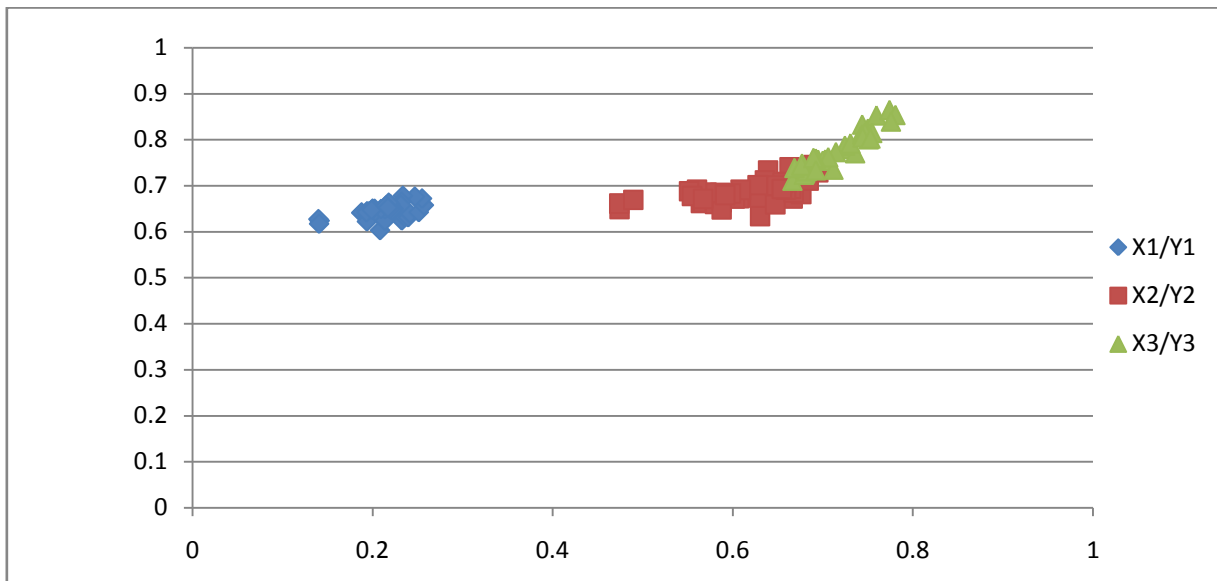


Fig 8: Graphical illustration of cluster center data

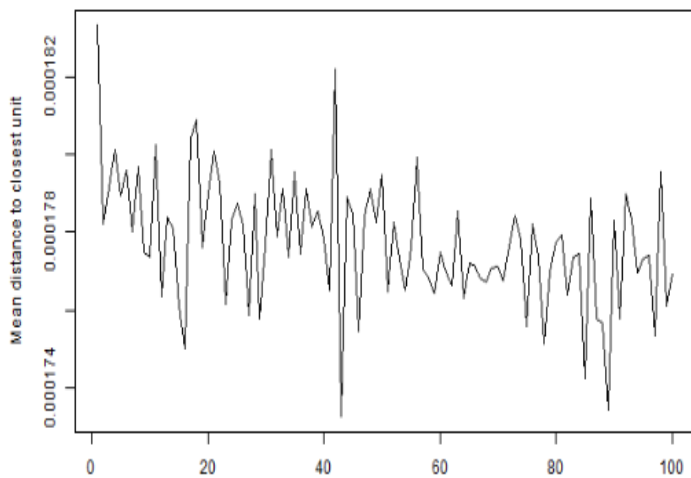


Fig 9: Training progress of SOM

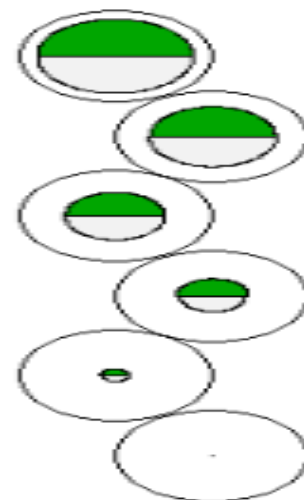


Fig 10: Plot of learning SOM



Fig 10: Class count plot of trained SOM

## 8. CONCLUSION

The research work made an attempt to develop an intelligent system for soil classification using unsupervised learning approach. The system was able to classify the soil into various classes. This is useful in many cases when it is necessary to know the type of soil. For example in farming, for customized application of fertilizer or matching of crop types to farmland. The system can be developed further for application in the mentioned areas.

## 9. REFERENCES

- [1] Agris Nikitenko (2006). Hybrid Intelligent System Development and Implementation. *Summary of Doctoral Thesis*. RIGA TECHNICAL UNIVERSITY, Riga.
- [2] Amit Konar (1999). Artificial Intelligence and Soft Computing: Behavioral and Cognitive Modeling of the Human Brain. Department of Electronics and Telecommunication Engineering, Jadavpur University, Calcutta, India. ISBN 0-8493-1385-6 (alk. paper), QA76.9.S63 K59,pg 28.
- [3] Baojuan Zheng (2008). Using Satellite Hyperspectral Imagery to Map Soil Organic Matter, Total Nitrogen and Total Phosphorus. Thesis Submitted to the faculty of the University Graduate School in partial fulfillment of the requirements for the degree Master of Science in the Department of Earth Sciences, Indiana University.
- [4] Cécillon, L.; Barthès, B.G.; Gomez, C.; Ertlen, D.; Genot, V.; Hedde, M.; Stevens, A.; Brun, J.J. Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS). *Eur. J. Soil Sci.* **2009**, *60*, 770–784.
- [5] Chang, C. W., D. A. Laird, M. J. Mausbach, and C. R. Hurburgh (2001). Near-Infrared Reflectance Spectroscopy—Principal Components Regression Analyses of Soil Properties. *Soil Sci. Soc. Am. J.*, *6*: 480–490.
- [6] Chen, F., D. E. Kissel, L. T. West, and W. Adkins (2000). Field-Scale Mapping of Surface Soil Organic Carbon Using Remotely Sensed Imagery. *Soil Sci. Soc. Am. J.*, *64*: 746–753.
- [7] Dalal, R. C. and R. J. Henry (1986). Simultaneous Determination of Moisture, Organic Carbon, and Total Nitrogen by Near Infrared Reflectance Spectrophotometry. *Soil Sci. Soc. Am. J.*, *50*: 120–123.
- [8] Dematte, J. A. M., M. V. Galdos, R. V. Guimaraes, A. M. Genu, M. R. Nanni, and J. Zullo (2007). Quantification of tropical soil attributes from ETM+/LANDSAT-7 data. *Int. J. Remote Sens.*, *28*(17): 3813–3829.
- [9] ESBN (European Soil Bureau Network), 2005. Soil Atlas of Europe. European Commission, Office for Official Publications of the European Communities, L-2995 Luxembourg.
- [10] He, Y., M. Huang, A. Garcia, A. Hernandez, and H. Song (2007). Prediction of soil macronutrients content using near-infrared spectroscopy. *Comput. Electron. Agr.*, *58*(2): 144–153.
- [11] Machine Learning, Part I: Supervised and Unsupervised Learning. [www.aihorizon.com/essays/generalai](http://www.aihorizon.com/essays/generalai)
- [12] Matlin, W. Margaret, *Cognition*, Hault Sounders, printed and circulated by Prism books, India, 1996.
- [13] Muller A.J. and S. Nilsson (2009). Harmonized World Soil Database (version 1.1), Foreword to FAO/IIASA/ISRIC/ISS-CAS/JRC. FAO, Rome, Italy and IIASA, Laxenburg, Austria.
- [14] Nikola Kasabov (2007). Brain, Gene and Quantum inspired Computational Intelligence: Challenges and Opportunities. *Challenges for Computational Intelligence*. DOI 10.1007/978-3-540-71984-7\_9, Vol 63, pp193–219
- [15] Ray S.S., J. P. Singh, G. Das, S. Panigrahy (2004). Use of High Resolution Remote Sensing Data for Generating Site-specific Soil Mangement Plan. XX ISPRS Congress, Commission 7. Istanbul, Turkey The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences: 127–131.
- [16] Samet, H., “Neighbor finding techniques for images represented by quadtrees,” *Computer, Graphics and Image Processing*, vol. 18, pp. 37–57, 1982.
- [17] Sergio Freire, Ines Fonseca, Ricardo Brasil, Jorge Rocha, and Jose A. Tenedorio (2013). Using Artificial Neural Networks for Digital Soil Mapping a comparison of MLP and SOM approaches. AGILE 2013 – Leuven, May 14–17.
- [18] Stephan Gmur, Daniel Vogt, Darlene Zabowski and L. Monika Moskal(2012). Hyperspectral Analysis of Soil Nitrogen, Carbon, Carbonate, and Organic Matter Using Regression Trees. *Sensors*. ISSN 1424-8220, doi:10.3390/s120810639, pg 10639–10658. [www.mdpi.com/journal/sensors](http://www.mdpi.com/journal/sensors)
- [19] Udelhoven, T., C. Emmerling, and T. Jarmer (2003). Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. *Plant Soil*, *251*(2): 319–329.
- [20] Vandal Smith (2010). Application of Neural Network in Whether Prediction. “*The Pacesetter* “. Vol. 10, No 3, 25–35.
- [21] Vogt, K.A.; Patel-Weynand, T.; Shelton, M.; Vogt, D.J.; Gordon, J.C.; Mukumoto, C.; Suntana, A.S.; Roads, P.A. *Sustainability Unpacked: Food, Energy and Water for Resilient Environments and Societies*; Earthscan: London, UK, 2010.
- [22] Zhengyong Zhaoa, Thien Lien Chowb, Herb W. Reesb, Qi Yanga, Zisheng Xingb and Fan-Rui Menga (2009). Predict Soil Texture Distributions using an Artificial Neural Network Model. *Computers and Electronics in Agriculture*. Pg 36–48. [www.elsevier.com/locate/compag](http://www.elsevier.com/locate/compag).
- [23] ([www.aihorizon.com/essays/generalai](http://www.aihorizon.com/essays/generalai)).