# Increasing Crop Production Based on Data Mining Concepts

Revathi S<sup>1</sup>, Akash P<sup>2</sup>, KarthiSankar M<sup>3</sup>, B.Keerthana<sup>4</sup>, S.Ashwini<sup>5</sup> <sup>1</sup>Assistant Professor, <sup>2,3 4 5</sup>UG Student <sup>1,2,3,4,5</sup> Department of CSE <sup>1,2,3,4,5</sup> Sri Ramakrishna Institute of Technology Coimbatore, India

## Abstract

Effective crop management is based on selecting suitable crops based on soil properties in certain region. Data mining algorithms and their application plays a vital role in the domain of agriculture. Most of the crops are preferred based on the physical and chemical properties of soil to achieve the successful crop management which leads to the higher yield performance. This work surveys and analyzes soil properties in existing research with respect to the following areas: 1. Main objective of the research, 2.Soil parameters, 3.Data mining technique or algorithm, 4. Performance analysis of existing works. The main focus is on the application of data mining techniques in agricultural field and also the process of discovering the knowledge in from soil database. It also explores the various chemical and physical properties of soil in agricultural land.

**Keywords** - Data mining techniques, knowledge discovery, Crop management, higher yield, Soil properties

# I. INTRODUCTION

Data Mining plays a major role in research works related to the domain of agriculture. Data Mining involves the process of extracting hidden information which are used for prediction, from large databases. Prediction plays an important role in covering different breadths of agriculture. Soil profile is one of the most important component that decides the growth and yield of crops. The yield obtained from a crop can be obtained through prediction, for which mining of large databases play a hidden role. The major challenge is the extraction of knowledge from vast amount of data. This process is known as knowledge discovery, in data mining. Various data mining techniques such as Clustering, Classification, Regression and Association can be used to analyze large dataset of soil related attributes. In this survey, the various soil attributes are identified and analyzed by various researchers based on certain diverse areas. Data mining techniques can be used to identify site specific crop management. This work provides a potential benefit to the farmers and is also a valuable tool to study the soils. Various analysis methods such as ANNOVA (Analysis of Variance), PCA (Principle Component Analysis) are used by the various researchers for prediction. The major research

challenge is the heavily diverse parameters associated with the soil at each geographic location.

#### **II. RELATED WORKS**

Shakil et al [1] proposed that the knowledge extracted from agricultural data can be used to estimate major cereal crop yields in major districts of Bangladesh, using the data mining

techniques. The soil parameters taken into account for this study are as follows. Maximum pH of soil, minimum pH of soil, soil salinity and other environmental variables such as rainfall, humidity, maximum temperature, average sunshine were considered on this research. Two parameters namely irrigated area and cultivated area were considered under the area central input attributes. Various clustering techniques were applied over these attributes, followed by suitable classification techniques to predict crop yield. The classification and regression models in Rapid Miner were used to select the input variables which were used for obtaining the prediction results.

Wang et al [2] suggested the derivation of Spatiotemporal yield trend map of the alluvial plains located in South West Taiwan which grows paddy rice, using satellite images. The potential yield which are dependent on the soil factors were identified by the general soil survey data in order to relate the soil characteristics to yield classes in the Spatio-temporal yield trend map. Classification and Regression Tree analysis (CART) was used to perform comparison of soil characteristics between those consistently high yield areas and low yield areas. Parameters such as soil texture, profile, internal drainage, soil acidity were taken into account for this research.

Rossel et al [3] compared different data mining algorithms for modeling soil visible-near IR diffuse reflectance spectra. These algorithms were also used in determining the interpretability of the result. Soft Computing techniques such as Multiple Linear Regression (MLR), Partial Least Square Regression (PLAR), Multi-variate Adaptive Regression Splines (MARS), Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN) were used to determine the soil organic carbon, clay content and pH. The Principle Component Analysis (PCA) was used to compress the spectra. The resultant thus obtained was treated with the K-means clustering algorithm to perform classification of the different types of spectra in spectral library.

Borman [4] stated that the climatic change has an impact over the soil moisture which has an indirect impact over the soil texture. Virtual columns were introduced which has been used to assess the soil texture. In this research, the author has used a scheme known as soil vegetation atmosphere transfer scheme (SVAT). The SVAT was applied to the virtual columns to assess the soil texture. 31 soil texture classes had been gathered from five different climatic regions of Germany.

Long-term simulations were carried out in different soil texture, to identify the relationship between the soil moisture and soil texture. After inferring through the simulation results, indications were made to represent soil moisture and was found to differ from one region to another region based on the soil texture. Loam, sand, clay and silt are those soil textures that are taken into account.

Cecile et al [5] mainly focused on mapping properties of soil, similar to those stated by other researchers, such as organic carbon, cation exchange capacity (CEC),soil pH, calcium carbonate, free iron, soil texture such as clay, silt, sand by using vis-NIR (Visible near infra red). In this study 95 soil samples were considered in 192 soil fields within the 24.6 km2 area. HyMapspectra was used to predict the soil properties. This prediction was computed for those eight properties of soil which using the Partial Least Square Regression (PLSR) algorithm. The result successfully predicted the soil properties using digital soil mapping with the help of the hyper spectral data.

Bateni et al [6] worked on hydro-climatology, a technique which deals with the land surface hydrology that requires an accurate measure of soil thermal properties. This study found the various factors for the correlations to estimate soil thermal properties and also various parameters for thermal properties such as size of soil particle, content of minerals, porosity and content of water. These parameters are not readily available with the laboratory and so the measurement has to be taken from field experiment. In order to overcome this problem the author suggested that following algorithms such as Genetic Algorithm-Finite Difference (GA-FD) and hybrid numerical methods to evaluate the thermal properties of the soil. Results obtained from this study clearly interpret the thermal properties of soil which has been accurately estimated by the hybrid numerical method and thermal conductivity of soil.

Johan et al [7] studied the effects of soil properties on crop yield, for which different tillage were used. The experimentation also analyzed the relationship between farming depth, crop production, soil type and moisture on crop yield. Different soil textures such as clay, silt and sand were

taken for observation. Hydraulic conductivity was analyzed using the procedural ANNOVA. This research interpreted that the different tillage depth had somewhat decreased the yield performance on the heavier soils. This research recommended that a shallow working depth can be used for heavier soils. It also showed that the increment of tillage depth sometimes produces a negative result over the crop growth.

da Silva et al [8] recommended the maintenance and improvement of soil quality based on the effective management of soil and crop. It is because, the soil and crop has a direct influence in its structure and micro biota habitats. This study is focused on quantifying and qualifying the Homogeneous Morphological Units (HMUs). Three types of tillage treatments were done during the course of this study. They are

- (i) no-tillage planting system (NT),
- (ii) no-tillage with chiseling(NTC) and
- (iii) conventional tillage(CT).

Various analysis techniques used in this study includes De- trended Correspondence Analysis (DCA), Principle Component Analysis (PCA) and Multivariate Analysis. Finally, the research was concluded that superior quality of soil was found with the NT and NTC planting system.

Holzman et al [9] focused on estimation of crop yield in certain region. The Temperature Vegetation Dry Index (TVDI) technique was used to estimate the crop yield. The wheat-crop was taken into account in this study. MODIS-Aqua sensor was used to capture the soil moisture. Different techniques and parameters were used to estimate the crop yield. They include Enhanced Vegetation Index (EVI), Land Surface Temperature (LSTmin, LSTmax) and Moderate resolution imaging spectro radio meter. Results of this research were better compared to that of other researches. It was because those results obtained were found to be 68-79% and 76-81% of the yield variability for soya bean and wheat respectively. Thereby it concluded that the satellite data can provide better results for decision makers.

Nassi et al [10] conducted their experimentation in central Italy. Two fields were taken into account for this research to investigate the soil texture, irrigation and nitrogen fertilization on the yield of switch grass (*Panicum virgatum L.*) cultivations. Various soil parameters taken into account for this study were sand,

silt, clay, organic matter, pH, total nitrogen, available phosphorus, exchangeable potassium, field capacity, wilting point. In this experiment dry yield of crops were monitored from the first year to fourth year. The cultivators identified that some factor is affecting the crop yield. They also identified switch grass able to achieve good yields under the limited conditions. The methods used in this research are Statistical analysis, Rtool, Linear mixed model and LMER convenience functions. However the final result suggested that the major role during the process is the selection of a suitable cultivar for implementing energy chains.

Sara et al [11] considered two adjacent experimental fields in the Mediterranean environment which were based on different climatic conditions to cover crop mulching and to predict the biological and chemical responses of soil.

Conventional management was taken into account for this research. To assess the effects of mulching on soil quality, different soil samples were collected after the transplanting the tomato yield. The various chemical and physical properties such as soil carbon, soil nitrogen, C/N ratio, silt, sand, carbonates and other textures were measured in this experiment. PCA was also done with the JMP 9.0 statistical software package. The result identified in this study was that the interaction between cover crop species and climatic condition causes the mulching on soil quality and microbial function.

# **III. COMPARISON**

The below comparison outlines the various parameters used and the yield obtained as a result of the research.

TABLE I. PARAMETRIC COMPARISON

Researc	Concentration		
her	Parameters	Techniques	Crop
Shakil et al	pH of soil, minimum soi pH of soil, l salinity, rainfall, humidity, maximum temperature, average sunshine	Regression Rapid Miner (tool used)	Cereal
Wang et al	Spatio-temporal	Classification	
	yield trend map	and Regression Tree analysis	Paddy
		Multiple Linear Regression (MLR), Partial Least Square Regression (PLAR), Multi-	

Rossel et al	organic soil carbon, clay content and pH	variate Adaptive Regression Splines (MARS), Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN), Principal Component Analysis (PCA)	-
Borman et al	Soil texture	soil vegetation atmosphere transfer	-
Cecile et al	organic carbon, cation exchange (CEC),soi capacity 1 pH, calcium carbonate, free iron, soil texture	Partial Least Square Regression	-
Bateni et al	size of soil particle, content of minerals, porosity and content of water	Genetic Algorithm- Finite Difference	-
Johan et al	cro farming depth, p production, soil type and moisture	ANNOVA	

Researcher	Concentration		
	Parameters	Techniques	Crop
Da Silva et al	Tillage Treatment	De-trended Correspondence Analysis (DCA), Principle Component Analysis (PCA)	-
l		and Multivariate Analysis	
Holzman et al	Temperature Vegetation Dry Index	Enhanced Vegetation Index (EVI), Land Surface Temperature (LSTmin, LSTmax) and Moderate resolution imaging spectro radio meter	-
Nassi et al	clay sand, silt, , organic matter, pH, total nitrogen, available phosphorus,	Statistical analysis, R-tool, Linear mixed model and	Switch grass (Panicum

	exchangeable	LMER	virgatum
	fiel potassium, d capacity, wilting point	convenience functions	<i>L</i> .)
Sara et al	soil carbon, soil nitrogen, C/N ratio, silt, sand, carbonates and other textures	Principal Component Analysis	-

# **IV. CONCLUSION**

The analysis of the physical and chemical characteristics of soil and also the crop growth status can vary from one place to another place. It was also identified that the yield restricted soil factors with the help of data mining technique. The yield level of the crop can also depend upon the soil characteristics because the soil is the prime resource to provide the required nutrients for crop growth. This survey provides an overall ideas on the various soil properties and the different ways in which crop growth can be managed using suitable data mining techniques.

## REFERENCES

- [1] Shakil Ahamed, A.T.M.; Mahmood, N.T.; Hossain, N.; Kabir, M.T.; Das, K.; Rahman, F.; Rahman, R.M., "Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh," in Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2015 16th IEEE/ACIS International Conference on , vol., no., pp.1-6, 1-3 June 2015 doi: 10.1109/SNPD.2015.7176185
- [2] Yi-Ping Wang. Yuvan Shen. "Identifying and characterizing yield limiting soil factors with the aid of remote sensing and data mining techniques", Precision Agriculture, 2015, Vol. 16(1), pp. 99-118 doi: 10.1007/s11119-014-9365-6
- [3] R.A.Viscarra Rossel, T.Behrens. "Using data mining to model

and interpret soil diffuse reflectance spectra", Geoderma, 2010, Vol. 158, pp. 46-54 doi: 10.1016/j.geoderma.2009.12.025

- [4] H.Borman, "Assessing the soil texture –specific sensitivity of simulated soil moisture to projected climate change by SVAT modeling", Geoderma, 2012, Vol. 185-186, pp. 73-83 doi: 10.1016/j.geoderma.2012.03.021
- [5] Cecile Gomez, Phillipe Lagacherce, Guillaume Coulouma, "Regional predictions of eight common soil properties and their spatial structures from hyper spectral Vis-NIR data", Geoderma, 2012, Vol. 189-190, pp.176-185 doi: 10.1016/j.geoderma.2012.05.023
- [6] S.M. Bateni, D.-S. Jeng, S.M. Mortazavi Naeini, "Estimating soil thermal properties from sequences of land surface temperature using hybrid genetic algorithm-finite difference method", Engineering Applications of Artificial Intelligence, 2012, Vol. 25, pp. 1425-1436 doi: 10.1016/j.engappai.2012.02.017
- [7] Johan Arvidsson, Aron Westlin, Fredrik Sörensson, "Working depth in non-inversion tillage-effects on soil physical properties and crop yield in Swedish field experiments", Soil and Tillage Research, 2013, Vol. 126, pp. 259-266. doi: 10.1016/j.still.2012.08.010
- [8] Adriana Pereira da Silva, Adriana Pereira, Letícia Carlos Babujia, Julio Cezar Franchini, Ricardo Ralisch, Mariangela Hungria, and Maria de Fátima Guimarães. "Soil structure and its influence on microbial biomass in different soil and crop management systems." Soil and Tillage Research, 2014, Vol. 142, pp. 42-53 doi: 10.1016/j.still.2014.04.006
  [9] M.E.Holzman, R. Rivas, M.C. Piccolo, "Estimating soil
- [9] M.E.Holzman, R. Rivas, M.C. Piccolo, "Estimating soil moisture and the relationship with crop yield using surface temperature and vegetation index" International Journal of Applied Earth Observation and Geoinformation, 2014, Vol. 28, pp.181-192 doi: 10.1016/j.jag.2013.12.006
  [10] N.Nassi o Di Nasso, M.V. Lasorella, N. Roncucci, E. Bonari, "
- [10] N.Nassi o Di Nasso, M.V. Lasorella, N. Roncucci, E. Bonari, " Soil texture and crop management affect switchgrass (Panicum virgatum L.) productivity in the Mediterranean", Industrial Crops and Products 2015, Vol. 65, pp. 21-26 doi: 10.1016/j.indcrop.2014.11.017
- [11] Sara Marinari, Roberto Mancinelli, Paola Brunetti, Enio Campiglia, "Soil quality, microbial functions and tomato yield under cover crop mulching in the Mediterranean environment", Soil and Tillage Research , 2015, Vol. 145, pp. 20-28 doi: 10.1016/j.still.2014.08.002