HAL
open science

Using parallel random forest classifier in predicting land suitability for crop production

Kennedy Senagi, Nicolas Jouandeau, Peter Kamoni

To cite this version:

Kennedy Senagi, Nicolas Jouandeau, Peter Kamoni. Using parallel random forest classifier in predicting land suitability for crop production. Journal of Agricultural Informatics, 2017, 8 (3), pp.23-32. 10.17700/jai.2017.8.3.390 . hal-02317289

HAL Id: hal-02317289

https://hal.science/hal-02317289v1

Submitted on 15 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hungarian Association of Agricultural Informatics
European Federation for Information Technology in Agriculture, Food and the Environment

Journal of Agricultural Informatics. Vol. 8, No. 3
journal.magisz.org

Using parallel random forest classifier in predicting land suitability for crop production

Kennedy Senagi[1], Nicolas Jouandeau[2], Peter Kamoni[3]

I N F O
Received 17 Nov 2017
Accepted 4 Nov 2017
Available on-line 21 Nov 2017
Responsible Editor: M. Herdon

Keywords:
machine learning, parallel random forest, land evaluation, soil analysis

A B S T R A C T

In this paper, we present an optimized Machine Learning (ML) algorithm for predicting land suitability for crop (sorghum) production, given soil properties information. We set-up experiments using Parallel Random Forest (PRF), Linear Regression (LR), Linear Discriminant Analysis (LDA), KNN, Gaussian Naïve Bayesian (GNB) and Support Vector Machine (SVM). Experiments were evaluated using 10 cross fold validation. We observed that, parallel random forest had a better accuracy of 0.96 and time of execution of 1.7 sec. Agriculture is the main stream of food security. Kenya relies on agriculture to feed its population. Land evaluation gives potential of land use, in this case for crop production. In the Department of Soil Survey in Kenya Agriculture and Livestock Research Organization (KALRO) and other soil research organizations, land evaluation is done manually, is stressful, takes a long time and is prone to human errors. This research outcomes can save time and improve accuracy in land evaluation process. We can also be able to predict land suitability for crop production from soil properties information without intervention of a soil scientist expert. Therefore, agricultural stakeholders will be able to efficiently make informed decisions for optimal crop production and soil management.

1. Introduction

One of the anchor pillars of Kenya's Vision 2030 is agriculture. Agriculture is an important sector in Kenya for economic growth and safe food supply (Government of Kenya 2007). Almost a quarter of Kenya Gross Domestic Product is from agriculture. Kenya sets aside a significant amount of money each year in agriculture through the state Department of Agriculture (in Ministry of Agriculture Livestock and Fisheries) whose mandate is: to promote and facilitate production of food and agricultural raw materials to enhance safe food supply and incomes, build capacity of agro-based (Ministry of Agriculture, Fisheries and Livestock 2016).

Due to low yields, change in farming practices and other factors, land evaluation has become an important practice for agricultural stakeholders. Land evaluation is the assessment of use potential of land. Land evaluation involves the analysis and interpretation of climate, vegetation, soil, socio-economic and other aspects of land such as crop land use requirements. Land evaluation predicts expected suitability of land use (e.g. crop production) for each land mapping unit. Ideally, representative soil profiles (land mapping units) are sampled horizontally and depth. The soil samples are then taken to the laboratory where the soil physical and chemical compositions are analyzed to determine the land qualities (e.g. soil water holding capacity, rooting depth, salinity hazard and oxygen availability). Besides the land qualities, economic factors (e.g. cost of inputs,

[1] Kennedy Senagi
Dedan Kimathi University of Technology and University of Paris 8
kennedy.senagi@dkut.ac.ke
[2] Nicolas Jouandeau
University of Paris 8
n@ai.univ-paris8.fr
[3] Peter Kamoni
Kenya Agriculture and Livestock Research Organization
pkamoni@gmail.com

price of benefits, and net returns) and climate (e.g. temperature and rainfall) characteristics are determined and rated against a scheme.

Thereafter, crop requirements are matched against the rated scheme to determining the suitability of a crop. According to Food Organization Authority (FAO) of the United Nations, suitability of a land mapping unit is categorized as either S1 (highly suitable), S2 (moderately suitable), S3 (marginally suitable) or NS (unsuitable). With this land evaluation information, farmers gain knowledge of what crops to grow in a particular piece of land, this leads to better crop yields as well as acquiring skills for sustainable land/soil management (Food and Agriculture Organization 1976).

Sorghum is a significant crop in improving food security in Kenyan (Ogeto, Cheruiyot, Mshenga & Onyari 2013). It is a drought tolerant crop; it can grow well in Arid and Semi Arid Lands (ASAL) in Kenya. ASAL make up-to 83% of Kenya (Dietz, Foeken, Soeters & Klaver 2014). Sorghum production is increasing because of the increased demand by industries (e.g. East African Breweries) and the market (Ogeto, Cheruiyot, Mshenga & Onyari 2013).

Machine Learning (ML) is a discipline in computer science that emerged from the field of artificial intelligence. ML is concerned with developing computer programs that automatically learn and improve with experience. To build an intelligent machine, it should learn from and adapt to the environment it learns in (Mitchell 1997). This research we compare performance of PRF, LR, LDA, KNN, GNB and SVM. Experiments show that PRF had a better performance in predicting land suitability given unseen soil profile data. This research can be adopted to improve efficiency of land evaluation process for crop production and soil management. In this paper, Section 2 gives related works, Section 3 covers methodology, Section 4 presents results and discussion and Section 5 gives conclusion.

2. Related Works

Anitha and Acharjya (2017) studies aimed at predicting crops to be grown in Vellore District. They came up with a hybridizing of rough set on fuzzy approximation space and ANN. The rough set on fuzzy approximation space was to get almost equivalence classes where attribute values are not qualitative. The classified information was the taken to ANN algorithm for training, prediction and testing. They used a dataset collected from Krishi Vigyan Kendra of Vellore District between 2007 and 2014. The dataset contained 2193 objects with 15 attributes. They used 26 soil attributes: soil pH, moisture, organic matter, availability of nitrogen, availability of phosphorus, availability of potassium, water pH, calcium, nitrate, magnesium, selenium rainfall, copper, zinc, manganese and iron. The dataset was divided into 55% training and 45% testing data and validated by N-folds cross-validation. The experiments were developed in R language. The average Mean Square Error (MSE) was 0.2436 and Accuracy of 93.2% (Anitha & Acharyjya 2017).

Fereydoon et al. (2014) developed an SVM – Two Class Model for land suitability analysis for wheat production in Kouhin region in Iran. They used 22 soil representative soil profiles information, each soil profile having 10 features: climatic (precipitation, temperature), topographic (relief and slope) and soil-related (texture, $CaCO_3$, Organic Carbon, coarse fragment, pH, Gypsum). They implemented a Two Class SVM model on a non-linear class boundaries i.e. non-linear mapping of input vector into a high dimensional feature space. They used MATLAB 8.2 software to design and test the SVM model. They randomized the dataset and split it into training (80%) and test (20%). In performance evaluation, they got an RMSE of 3.72 and R^2 of 0.84 (Fereydoon et al. 2014).

Land suitability classification using a large number of parameters is time consuming and costly. With this research problem Hamzeh et al. (2016) presented a combination of feature selection (best search, random search and genetic search methods) and fuzzy-analytical hierarchical process (AHP) methods to improve selection of important features from a large number of parameters. On feature selection, random search performed slightly better than genetic search methods and best search. They found that soil texture, wetness, salinity and alkalinity were the most effective parameters for determining land suitability classification for the cultivation of barely in the Shavur Plain,

southwest Iran. The report showed that soil salinity and alkalinity, soil wetness, $CaCO_3$, gypsum, pH, soil texture, soil depth and topography were the most important soil properties to consider for cultivating barley in the study area (Hamzeh et al. 2016).

Mokarram et al. (2015) used AI and ML to automate the land suitability classification for growing wheat. They used a dataset with data collected from Shavur plain, northern of Khuzestan province, southwest of Iran. The dataset had the following attributes: topography (Primary slope, Secondly slope and Micro relief), salinity and alkalinity (EC and ESP), wetness (Groundwater depthand Coroma depth), soil texture, $CaCO_3$, soil depth, Gypsum and pH (H2O). Land suitability classes were classified as highly suitable (more than 75%), moderately suitable (50%-75%), marginally suitable (25%-50%) and not suitable (less than 25%). Mokarram et al. (2015) implemented Bagging, AdaBoost and RotForest algorithms and evaluated the performance of their experiments using 632+ bootstrap and 10-fold cross validation. Their results classified 26% of the land being moderately suitable, 25% being marginally suitable and 49% being not suitable. Moreover, they found that RotBoost algorithm had a better accuracy than Single Tree, Rotation Forest, AdaBoost, Bagging algorithms in predicting land suitability class. RotBoost recorded 99% and 85% accuracy score for bootstrap and cross validation respectively (Mokarram, Hamzeh, Aminzadeh & Zarei 2015).

Dahikar and Rode (2014) demonstrated the use of ANN in predicting crop suitability from soil attributes. The attributes include: type of soil, pH, nitrogen, phosphate, potassium, organic carbon, sulphur, manganese, copper, calcium, magnesium, iron, depth, temperature, rainfall, humidity. They set up the experiments in MatLab (Dahikar & Rode 2014). However, performance results were not given in the papers. The researchers acknowledge the potential of using ANN in predicting crop suitability from soil data collected from rural district.

Elsheikh et al. (2013) presented ALSE, which was an intelligent system for assessing land suitability for different crops (e.g. mango, banana, papaya, citrus, and guava) in tropical and subtropical regions based on geo-environmental factors. ALSE supported GIS capabilities on the digital map of an area with FAO-SYS framework model. It also had some necessary modifications to suit the local environmental conditions for land evaluation, and the support of expert knowledge. ALSE had the capability to identify crop-specific conditions and systematically computes the spatial and temporal data with maximum potential. This would help land planners to make complex decisions within a short period taking into account sustainability of a crop. Test dataset was collected from agricultural land in Terengganu, West Malaysia. Some of the attributes they used were: rainfall, soil attributes (nutrient availability, rooting conditions, nutrient retention, soil workability and oxygen soil drainage class) and topology (Elsheikh et al. 2013).

Manuel et al. (2014) evaluated 179 classifiers. They were implemented in Weka, Matlab, R (with and without the caret package) C and other relevant classifiers. PRF turned out to be the best classifier when implemented in R which achieved 94.1% accuracy. Followed by SVM with Gaussian and polynomial kernels which achieved 92.3% accuracy. Different learning algorithms had different performance levels on change in dimensionality (Manuel et al. 2014).

Rich and Alexandru (2006) did an empirical evaluation of supervised learning on high dimension data. They evaluated their experiments using: accuracy, area under the curve, and squared error. They also studied the effect of increasing dimensionality using SVM (LaSVM kernel and RBF kernels using stochastic gradient descent), ANN, Logistics Regression and Nave Bayes. They found that RF had the best performance, followed by ANN, then boosted trees, and lastly SVMs (Rich and Alexandru, 2006).

From the above literature, we see diverse applications of statistical and ML techniques in land suitability for better crop production: ANN, AdaBoost, SVM, fuzzy logic and PCA. Nevertheless, other researchers compared performance of algorithms. For example, Manuel et al. (2014) evaluated 179 classifiers including: discriminant analysis, multiple adaptive regression splines, Bayesian, rule-

based classifier, boosting, neural networks, bagging, stacking, Random Forest (RF) (Breiman 2001), SVM, decision trees, generalized linear models, logistic and multinomial regression, nearest neighbours, partial least squares and principal component regression. PRF turned out to have the best accuracy of 94.1% (Manuel, Eva & Senen 2014). Liua et al. (2013) did a comparison of back propagation neural network, SVM and RF for investigating electronic tongue (E-tongue) data processing. RF had a better performance of 99.07% (Liua, Wangb, Wanga & Li 2013). Palmer et al. (2007) implemented RF, Partial-Least-Squares regression and ANN to develop QSPR models for predicting aqueous solubility. RF had a better prediction accuracy (Palmer, O'Boyle, Glen & Mitchell 2007). Jia et al. (2013) researched on predicting β-hairpin motifs in proteins. They applied RF and SVM. RF had a better accuracy of 83.3% and Matthews correlation coefficient of 5-fold cross-validation of 0.59 (Jia, Hu & Sun 2013). Dong et al. (2016) found that RF outperformed SVM in a framework for eye state estimation with various feature sets in terms of time consumption (Dong, Zhang, Yue & Hu 2016). We see that PRF is a potential algorithm in optimizing land suitability for Sorghum production, from soil properties dataset.

3. Methodology

We set up experiments of PRF for predicting land suitability from soil datasets. We further compared of PRF predictability with other algorithms, namely, LR, LDA, KNN, GNB and SVM.

3.1. How Parallel Random Forest Works

RF is an ensemble ML algorithm. Ensembles use the divide-and-conquer approach when learning from a dataset. Growing ensembles of trees (by generating random vectors for each tree) and letting them vote for the most popular class has resulted to significant improvements in classification accuracy. RF applies a random split of each node in the tree from possible best splits. RF also does a random selection of a subset of features to use to grow each tree. RF generates new training sets by randomizing the outputs in the original training set or generating training set from a random set of weights on examples in the training set. The randomness minimizes correlation while maintaining strength. The main principle behind ensemble is that they use a group of weak learners to form a strong learner. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them (Breiman 2001). PRF is a version of RF which distributes computations of decision trees on multi-core architectures (Kulkarni & Sinha 2013).

3.2. Datasets and Experiments

This study identified eight features that can be used in land evaluation for Sorghum production. These features were: drainage, annual rainfall, rooting depth, salinity hazard (ECE), sodicity hazard (ESP), moisture holding capacity (MHC), slope and mean annual temperate. In land evaluation suitability for crop production, soil scientists consider crop requirements, land qualities rating schemes and soil properties information when evaluating the optimal soils for crop production. This research considered soil properties data from reports archived in KALRO. This data was selectively collected from Tsavo, Muhoroni, Bura East, Machakos, Olenkasoria, Sagana, Bura West, Kericho, Makuyu,Busia, Nakuru, Kiboko, Amboseli, Taveta, Kwale, Kapenguria, Kindaruma, and Chuka areas between 1977 to 1996.

We developed a Land Evaluation Software (LES). We followed the rapid prototyping procedures of software development using Microsoft Visual Studio 2013. Figure 1 is an interface of the LES. The following were functions of the LES:

a) Import dataset in the system from CSV files. The data include: crop requirements, soil rating criteria and soil properties data.

b) Analyse soil dataset to give land suitability using minimum principle. Soil limiting factors are also given.

c) Export analysed data to CSV file for users to make informed decisions on soil management and crop production.

In LES, the user customizes crop requirements in a csv file. The Import Data Crop Req. button allows the user to browse the file system, select the crops requirements csv file which will be loaded in the Crop Requirements data grid. Import Data (Soil Data) button loads soils data to be analysed from a csv file. After the crop requirements and the soil dataset is loaded as show in Figure 1, the use clicks the Analyse Suitability button. This button initiates analysis of the soil dataset as per the following land evaluation procedures (Food and Agriculture Organization 1976):

(i) An appropriate soil quality/feature is matched to a soil rating scheme to get *r*.

(ii) *r* is then matched to crop requirement, to get land suitability, *s*

(iii) An evaluated LMU have several properties. If all the properties are evaluated in steps (i) and (ii) above, they will give us $S=\{s_1,\ldots,s_n\}$, where n = *number of soil properties*.

(iv) Therefore, the overall land suitability is the minimum value, *s_min*, in the set *S*, as shown in equation 1

$$(1)\quad s_min = \min\{s_1, \ldots, s_n\}$$

(v) Soil limiting factors for each land mapping unit are recorded. Soil limiting factors are the soil properties that have recorded low suitability as per steps (i), (ii) and (iii).

The overall land suitability and limiting factors for each LMU are displayed in Overall Suitability and Limiting columns respectively in the Soil Analysis data grid in LES. Note that in this research, we did not consider soil fertility aspects in land evaluation. The steps (i) to (iv) described above are: done manually, time consuming, stressful and prone to human error, however, LES is accurate and fast. The output of this evaluation can be used by soil scientists to advise farmers on the land potential. Knowing the soil limiting factors assist soil scientists' advice of the appropriate chemicals or fertilizers to be added in the soil, to improve crop production thus improving soil management. LES can be adopted for land evaluation for crop production and soil management. LES can also be upgrade to store and management soil data and information for research purposes. Ultimately, farmer would be able to get timely and accurate results for optimal crop production. In general, LES is a resourceful software

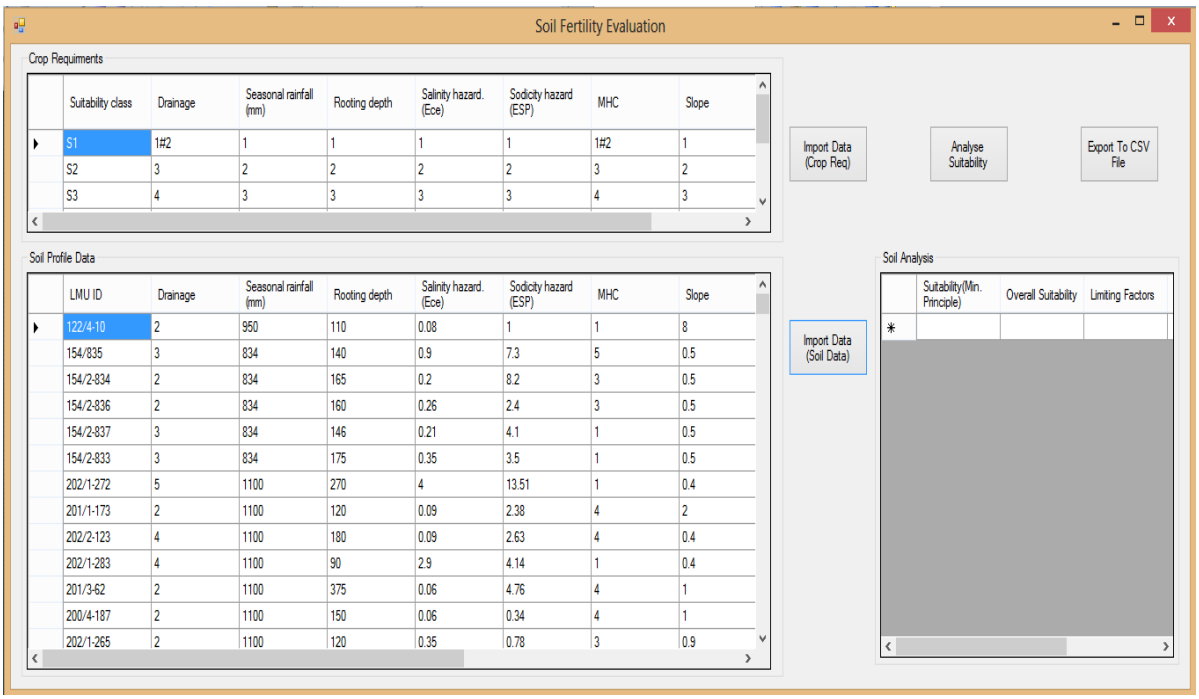

Figure 1. Land Evaluation Systems interface

In ML algorithms, after pre-treatment of the dataset in LES and exporting to csv file using the Export To CSV File button, the overall suitability is taken as the label while the soil properties are taken as the features. In ML, the exported dataset was split to a ratio of 7 to 3, that is, training and test set respectively. We trained, tested and measured predictability of PRF, LR, LDA, KNN, GNB and SVM. The ML algorithm performance was measured by 10 cross fold validations. We prototyped our experiments in Python 3.5 running on a computer with Intel Core i3-2365M @1.40GHz with 4GB RAM.

4. Results and Discussion

We visualized the distribution of soil features population as show in Figures 2 and 3. The horizontal axis in Figure 2 indicates levels of features in the soil while the vertical axis indicates the population. In Figure 3, vertical axis indicates the levels of features in the soil. Comparing Figure 2 and 3, we only see moisture holding capacity distribution can fit a gaussian distribution. The other features do not show a gaussian distribution. This can guide us in choosing an appropriate ML algorithm e.g GNB for gaussian distributed features. In Figures 2, we also see rooting depth, salinity hazard, slope and sodicity hazard populations highly skewed to the lower values. Same as in Figure 3, we see the inter-quartile ranges and upper whiskers flattened to the lower populations as the far flanked levels marked as outliers. In Figure 2, we see the median being same as the lower quartile. This is because there were many values drainage recorded at 3.0. This brought the median down.

In the soil dataset, we had 4 land suitability classes: highly, moderately, marginally and unsuitable. We trained, tested and measure performance of different algorithms: RF, LR, LDA, KNN, GNB and SVM ML algorithms. 10 cross validation was used in measuring performance of the ML algorithms. Results are shown in Table 1 and Figure 4.

Table 1 and Figure 4 shows 10 cross fold validation results. We see different ML algorithms adopting differently in the soil dataset. RF adopts quite well while SVM adopted quite poorly compared to the other ML algorithms. RF has a better accuracy (0.90 accuracy on 10 cross fold validation) and lower

standard deviation (0.13). This means RF correctly identified 90% of the unseen data correctly and a low variation from the mean of 0.13. In Figure 4, RF median is close to the upper bond quartile range and close to the upper whisker. This mean accuracy values in the upper median were quite high and had a low variation. Low standard deviation also shows that RF has stable results.

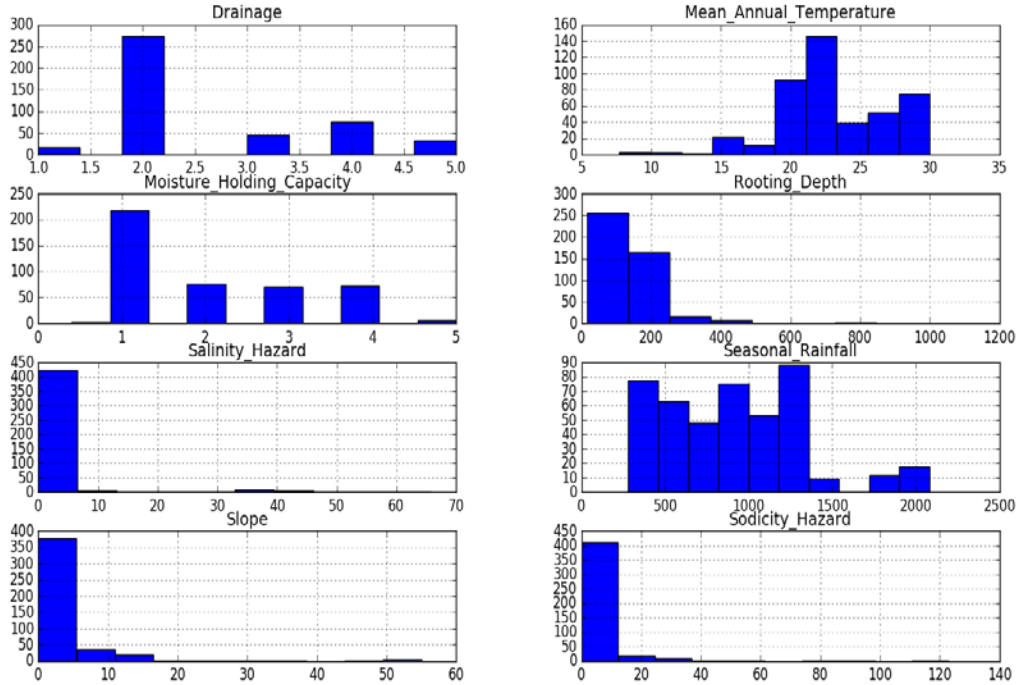

Figure 2. Bar graphs of population distribution concentration levels of different soil features in the dataset.

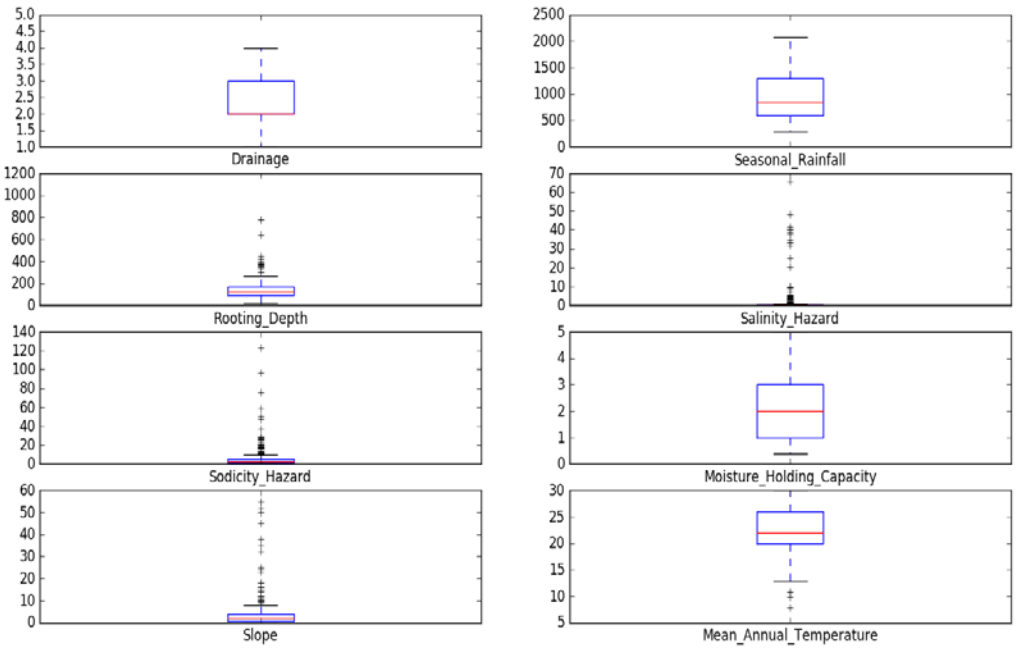

Figure 3. Box plot of population distribution in different soil features in the dataset.

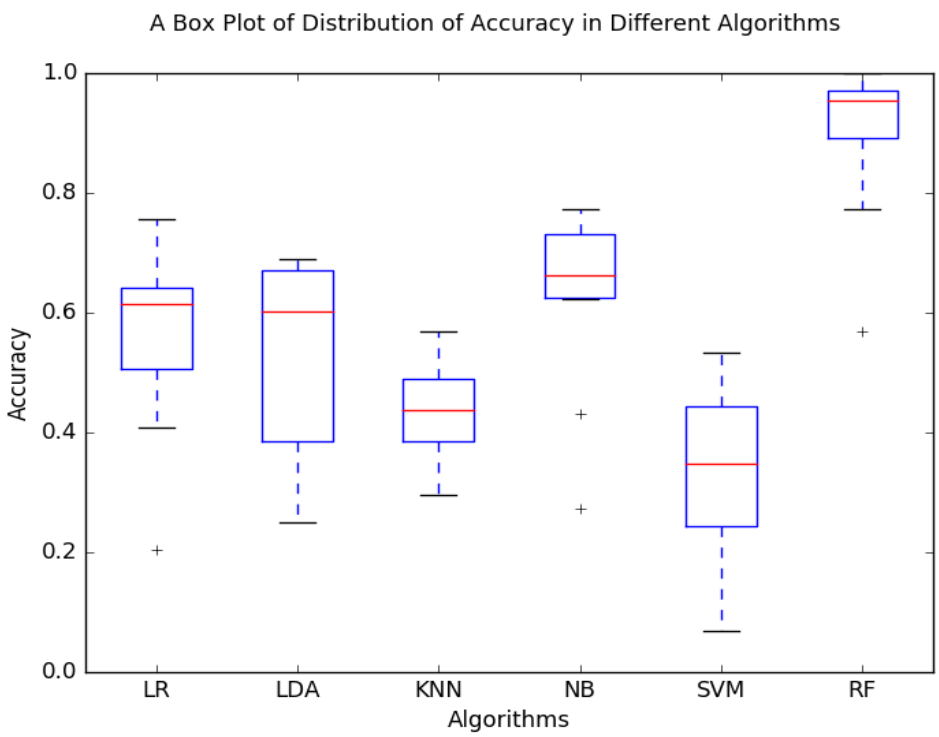

Figure 4. Distribution of accuracy in different machine learning algorithms applied on the soil dataset

Table 1. Cross validation acuracy and standard deviation of different ML algorithms

Algorithm	10 Cross Validation Accuracy	Standard Deviation
LR	0.56	0.16
LDA	0.53	0.16
KNN	0.44	0.08
Gaussian NB	0.63	0.15
SVM	0.34	0.14
RF	0.90	0.13

Results in Table 2 show serial and parallel RF. Accuracy was calculated using Equation 2. We see parallel RF runs 1.2 times faster than serial RF. Parallel runs faster because it utilizes multiple CPUs while serial RF runs on one CPU. More CPU increases processing capability of the computer by distributing processes across the CPUs making processes run faster (Rajaraman and Siva 2016). Moreover, accuracy does not change in parallelized and serial RF. Meaning increasing CPUs did not interfear with accuracy of RF.

Table 2. Comparing serial and parallel RF

Algorithm	Accuracy	Time (sec)	No. of CPUs
Parallel RF	0.96	1.7	4
Serial RF	0.96	2.1	1

$$(2)\quad accuracy\ (y, \hat{y}) = \frac{1}{n_{samples}}) \sum_{i=0}^{n_{samples}-1} 1 - (\hat{y} = y_i)$$

5. Conclusion

Land evaluation for crop production provides useful information to stakeholders in the agricultural sector to improve crop yield and soil management. In the Department of Soil Survey in KALRO and other soil research organizations, land evaluation is done manually, is slow and is prone to human errors. We developed LES that accurately and timely analyzed and labeled soil properties datasets. LES also indicated limiting factors/features in a set of soil properties; with this knowledge, soil scientists are in a better position to advise the appropriate chemical(s) or fertilizer(s) to add to improve crop production, thus improving soil management. Nonetheless, LES provided us with pre-treated labeled datasets that enabled us use PRF algorithm to predict land suitability for sorghum production from soil properties datasets. We evaluated the PRF experiments with actual datasets and recorded an accuracy and time of executing PRF of 0.96 and 1.7 sec respectively. PRF made accurate predictions without the involvement of a soil scientist human expert. Fertility can also be considered in land suitability evaluation by considering fertility aspects of soil profile information. In conclusion, LES can manage soil information while PRF can optimize prediction of land suitability for crop production from soil information. This will ultimately improve land evaluation process and provide agricultural stakeholders with appropriate information to improve soil management and enhance crop production.

References

Anitha, A & Acharyjya, D 2017. 'Crop suitability prediction in Vellore District using rough set on fuzzy approximation space and neural network', The Natural Computing Applications Forum, Springer. https://doi.org/10.1007/s00521-017-2948-1

Breiman, L 2001. 'Random forests', Machine learning, vol. 45, no. 1, pp. 5-32. https://doi.org/10.1023/A:1010933404324

Dahikar, S & Rode, S. 2014. 'Agricultural Crop Yield Prediction Using Artificial Neural Network Approach', International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, vol. 2, no. 1.

Dietz, T, Foeken, D, Soeters, S, & Klaver, W 2014. 'Agricultural dynamics and food security trends in Kenya', Developmental Regimes in Africa (DRA) Project ASCAFCA Collaborative Research Group: Agro-Food Clusters in Africa.

Dong, Y, Zhang, Y, Yue, J, & Hu, Z 2016. 'Comparison of random forest, random ferns and support vector machine for eye state classification', Multimedia Tools and Applications, Springer, vol. 75, no. 19, ISSN: 1380-7501, pp. 683-686.

Elsheikh, R, Shariff, A, Amiri, F, Ahmad, N, Balasundram, S & Soom, M 2013. 'Agriculture Land Suitability Evaluator (ALSE): A decision and planning support tool for tropical and subtropical crops', Journal of Computers and Electronics in Agriculture, Elsevier, vol. 93, pp. 98-110. https://doi.org/10.1016/j.compag.2013.02.003

Fereydoon, S, Ali, K, Azin, R, Ghavamuddin, Z, Hossein, J & Munawar, I 2014. 'Support Vector Machines Based-Modeling of Land Suitability Analysis for Rainfed Agriculture', Journal of Geosciences and Geomatics, vol. 2, no. 4, pp. 165-171.

Food and Agriculture Organization 1976. A framework for land evaluation. http://www.fao.org/docrep/X5310E/x5310e00.htm. FAO Soils Bulletin, No. 55. [June 2016]

Government of Kenya 2007. 'A Globally competitive and prosperous Kenya', Available from: www.opendata.go.ke. [November 2016]

Hamzeh, S, Mokarram, M, Haratian, A, Bartholomeus, H, Ligtenberg, A, & Bregt, A 2016. 'Feature Selection as a Time and Cost-Saving Approach for Land Suitability Classification (Case Study of Shavur Plain, Iran)', Journal of Agriculture, Multidisciplinary Digital Publishing Institute, vol. 6 no. 4.

Jia, S, Hu, X & Sun, L 2013. 'The Comparison between Random Forest and Support Vector Machine Algorithm for Predicting B-Hairpin Motifs in Proteins', Engineering, vol. 5, pp. 391-395.

https://doi.org/10.4236/eng.2013.510B079

Kulkarni, V & Sinha, P 2013. 'Random Forest Classifier: A Survey and Future Research Directions', International Journal of Advanced Computing, ISSN:2051-0845, vol. 36, no. 1, pp.1144-1153.

Liua, M, Wangb, M, Wanga, J & Li, D 2013. 'Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar', Elsevier.

Manuel, F, Eva, C, & Senen, B 2014. 'Do we Need Hundreds of Classifiers to Solve real World Classification Problems?', Journal of Machine Learning Research, pp. 3133-3181.

Ministry of Agriculture, Fisheries and Livestock 2016. Ministry of Agriculture, Fisheries and Livestock. Available from http://www.kilimo.go.ke/. [March 2017].

Mitchell, TM 1997. 'Machine Learning', ISBN: 0070428077, McGraw-Hill.

Mokarram, M, Hamzeh, S, Aminzadeh, F & Zarei, AR 2015. 'Using Machine Learning Land Suitability Classification', West African Journal of Applied Ecology, vol. 23, no. 1, pp. 63-73.

Ogeto, R, Cheruiyot, E, Mshenga, P & Onyari C 2013. 'Sorghum production for food security: A socio-economic analysis of sorghum production in Nakuru County, Kenya', Africa Journal of Agriculture and Research, pp. 6055-6067.

Palmer, DS, O'Boyle, NM, Glen, RC & Mitchell, JB 2007. 'Random forest models to predict aqueous solubility', Journal of chemical information and modeling, vol. 47, no. 1, pp. 150-158. https://doi.org/10.1021/ci060164k

Rajaraman V. and Siva R.M. 'Parallel Computers Architecture and Programming.' PHI Learning Pvt. Ltd, 2016

Rich C. and Alexandru N. M. 2006. 'An Empirical Comparison of Supervised Learning Algorithms,' Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, pp. 161-168.