Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE 7 (2020) 72–82

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

Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters

M.S. Suchithra *, Maya L. Pai

Department of Computer Science and IT, Amrita School of Arts and Sciences, Kochi, Amrita Vishwa Vidyapeetham, 682024, India

ARTICLE INFO

Article history:
Received 9 December 2018
Received in revised form
12 April 2019
Accepted 15 May 2019
Available online 21 May 2019

Keywords:
Extreme learning machine
Activation functions
Soil fertility indices
Soil pH
Classification
Prediction

ABSTRACT

The soil, Soul of Infinite Life, is the entity responsible for sustaining life on earth. In spite of significant advances in the service sector, agriculture remains the major provider of employment and source of revenue in India. Soil testing is a valuable tool for evaluating the available nutrient status of soil and helps to determine the proper amount of nutrients to be added to a given soil based on its fertility and crop needs. In the current study, the soil test report values are used to classify several significant soil features like village wise soil fertility indices of Available Phosphorus (P), Available Potassium (K), Organic Carbon (OC) and Boron (B), as well as the parameter Soil Reaction (pH). The classification and prediction of the village wise soil parameters aids in reducing wasteful expenditure on fertilizer inputs, increase profitability, save the time of chemical soil analysis experts, improves soil health and environmental quality. These five classification problems are solved using the fast learning classification technique known as Extreme Learning Machine (ELM) with different activation functions like gaussian radial basis, sine-squared, hyperbolic tangent, triangular basis, and hard limit. After the performance analysis of ELMs with diverse activation functions for these soil parameter classifications, the gaussian radial basis function attains the maximum performance for four out of five problems, which goes above 80% in most of the accuracy rate calculations in every problem, followed by hyperbolic tangent, hard limit, triangular basis, and sine-squared. However, the performance of the final classification problem, i.e. the pH classification, gives moderate values with the gaussian radial basis and best performance (near 90%), with the hyperbolic tangent.

1. Introduction

The major focus for soil management in agriculture for enhancing crop productivity is on the maintenance and improvement of dynamic soil parameters. The population stresses, terrestrial limitations and the decline of traditional soil management methods have directed to a deterioration in the fertility of the soil in developing countries like India. Crop health is a major element in the highly productive system of modern agriculture. A substantial increase in crop production can be attained by adopting the suitable crop health management strategy. The increased productivity could be

* Corresponding author.
E-mail address: suchithrams194@gmail.com (M.S. Suchithra).
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2019.05.003
2214-3173

achieved through effective soil resource management and corrective measures to apply micronutrients. Timely detection and controlling of problems connected with crop yield pointers enables the decision makers (agricultural experts) and farmers to decide on appropriate soil resource management and crop environment management. Nowadays the prediction and classification problems are effectively handled by Machine Learning (ML) techniques. The exposure of ML methods in the area of agriculture definitely reduces the challenges faced by the domain experts.

In earlier days of ML techniques, the Levenberg-Marquardt based back-propagation method in the Artificial neural networks (ANNs) was used to predict the soil fertility [1]. Partial least squares regression was also used to forecast the soil fertility using the available water capacity, electrical conductivity (EC), clay loam, silt loam, sandy loam soils, soil OC and soil bulk density [2]. Numerous studies have been applied with ML techniques to identify and solve soil problems in agriculture like the prediction of soil fertility, the supply of the required nutrient levels and water etc. [3]. The wheat yield was classified and predicted by J48, K-nearest neighbors (KNN), One-R and Apriori classifier methods using phenotypic plant traits as inputs [4]. Using supervised Kohonen and counter-propagation neural networks, the wheat yield has been quantified as low, medium and high [5]. Naive Bayes classifier, Decision Tree (DT), Random Forests (RF), Support Vector Machine (SVM), AdaBoost and Logistic Regression (LR) were used to make decisions about insecticide application for leafroller pest monitoring on kiwifruit [6]. An unbiased linear predictor was used to predict the soil organic carbon [7]. The organic carbon on Sicilian soils was predicted by using the boosted regression trees [8]. The random forest has been combined with the feature selection method of genetic algorithms to predict organic carbon on soils of eastern Australia [9]. Organic carbon alongside with soil acidity (pH) and Cation Exchange Capacity (CEC) from mid-infrared spectra for a number of soils were predicted by partial least squares [10]. The Bayesian network was applied to perform soil fertility rating with the pH value of soil and the soil nutrients like nitrogen, copper, iron, potassium, phosphorus, organic carbon and Zinc [11]. The geographic portability of machine learning algorithms has been evaluated for the forming of wind speed [12]. DT, RF, SVM, Bayesian Networks, and ANN methods were able to analyze soil and climate which are directly involved in precision farming and crop growth [13].

Different machine learning approaches were used to predict the soil nutrient content, soil type and soil moisture [14]. To categorize village wise soil fertility indices and soil nutrient levels, a collection of twenty classifiers including RF, AdaBoost, SVM, neural networks and bagging have been used and the class labels were calculated on a scale of low, medium and high based on their numeric values [15]. A wide collection of regression methods has been used to generate pedotransfer functions which straightly foretell the numeric values of village wise fertility indices [16]. The soil fertility information of India is summarized for district and block levels. This information is suitable for making decisions about proper quantity usage of fertilizers, the consumption in terms of variations in fertility levels and the procedure of fertilizer distribution.

The key objective of this research work is to classify area wise soil fertility indices based on the village level soil fertility information. This classification can be used in making a village wise fertility index analysis report and it would be used to make fertilizer recommendations with the decision support system. This report would facilitate a comparative study of levels of soil fertility among villages and hence the importance of a research work to classify the fertility indices for soil nutrients like OC, P, K and B. North Kerala is highly prone to soil erosion, leaching of soil nutrients, floods, and droughts; so predicting the type of soil pH can avoid the unnecessary usage of chemical fertilizers [17]. The interest in predicting the levels of these soil parameters with ML techniques helps in reducing the unnecessary spending on fertilizer inputs and analyses soil health and environmental quality.

This study aims to classify the soil fertility indices and pH levels of Kerala north central laterite region soil and to predict these values based on soil parameters. The soil test report values of corresponding soil parameters are used here for the experiment and analysis. The fertility indices are obtained by using the six-tier soil nutrients values i.e., Very Low (VL), Low (L), Medium (M), Moderately High (MH), High (H) and Very High (VH). Extreme Learning Machine (ELM) methods are one of the second generation neural network methods, which are suitable for classification problems as well as to get faster classification and prediction results. Here, different meta-parameters of ELM are tuned to find a better classification model for soil nutrients classification problem. The current study attempts to improve the accuracy rate of fertility indices classification of soil nutrients and soil pH level classification in Kerala agriculture using ELM methods. Finally, we focused on the neural network concepts for classification. Considering all the available parameters, neural network provides better prediction over statistical methods. The Extreme Learning Machine and its variants with different activation functions are precisely selected due to the worthy act that they demonstrated in the experimental evaluation [18]. This proposed ML technique is suitable for soil classification in Kerala with better accuracy than available methods. In the current agricultural scenario, an early prediction about the soil status using machine learning is very helpful for farmers to create a better environment for next cultivation.

2. Materials and methods

2.1. Study site

The geographic study region for the present work is the North Central Laterites Agro-Ecological Unit (AEU) in the State of Kerala, located at $74^0 52'$ and $77^0 22'$ East longitude and $8^0 18'$ and $12^0 48'$ North latitude with a total area of 38,864 km^2. The shift from the agrarian economy towards a service sector dominated economy is indicating that the agriculture in Kerala has undergone major structural variations in the form of a decline in the share of Gross State Domestic Product (GSDP). The share of GSDP has decreased from 55% during the 60′s to less than 10% in 2011–12. Yet agriculture remains a major source of income in terms of Kerala people's livelihood [19]. The agrarian scenario in Kerala is rather distinct and unique from other states of India in terms of terrestrial consumption and harvesting pattern. The soil test based advisory facilities and soil research in Kerala had started ten years back but their usefulness was not realized in terms of classification and prediction of soil nutrients to maintain the area based soil health status. The data used for this study is collected from the region of North Central Laterites by the State Government of Kerala in the country India during the years 2014 to 2017 [20] and the geographical study area is highlighted in Fig. 1. The climate of North Central Laterites is tropical humid monsoon type (mean annual temperature 27.6 ^{0}C; rainfall 2795 mm). The lowlands have strongly acid, non-gravelly clay soils with impeded drainage and the uplands have strongly acid, gravelly, lateritic, low activity, clay soils, often underlined by plinthite. The North Central Laterites region covers 171,469 ha (4.41%) in the State [17].

Samples of soil are collected from individual farmers by the soil testing laboratory. The area wise analysis results and the gradation details about each input measure are publically available [20,23]. The soil samples were analyzed for 11 parameters of immediate relevance to plant nutrition: soil reaction (pH), electrical conductivity (EC), OC, plant available primary nutrients (P, K), secondary nutrient (S) and micronutrients (Cu, Zn, Fe, Mn and B). The analytical methods used for estimating soil fertility parameters are: pH is estimated using a pH meter with 1: 2.5 soil water suspension, EC is a measure of the concentration of soluble salts and the extent of salinity in the soil is measured using a conductivity meter with 1:2.5 soil water suspension, OC is estimated with Walkley and Black's wet digestion method using potassium dichromate for oxidation of organic matter was used to estimate soil organic carbon [24], the extraction of available phosphorus in soils is done by using Bray No.1 reagent and the estimation of phosphorus is obtained by the ascorbic acid method [25], the extraction of available potassium in soils is prepared with soil: solution ratio of 1:5 of neutral normal ammonium acetate solution and the potassium in the extract is measured by flame photometry, available sulphur (S) in soils is extracted by 0.15% $CaCl_2$ solution with soil: solution ratio of 1:5 and shaking for 30 min and the filtered extract is measured for the content of Sulphur by turbidimetric procedure, available copper (Cu), Zinc (Zn), iron (Fe) and manganese (Mn) is extracted using 0.1 N HCL for acid soils and the elements in solution are estimated by atomic absorption spectrophotometry, available boron (B) in soils is extracted using Hot water Extraction procedure [26]. This is a multivariate study in which independent variables are pH, EC, OC, P, K, S, Cu, Zn, Fe, Mn and B, and dependent variables are soil fertility index levels of OC, P, K and B as well as pH levels.

2.2. *Classification of village wise soil fertility indices*

Incorrect soil and crop management practices during cultivation have given rise to a heavy loss in soil quality [27]. Chemical fertilizers used in excess have created imbalances with the availability of soil nutrients. The practices like liming which can rectify acidity are not given enough attention. The productivity of Kerala soils is affected by all these factors. As a consequence of specific soil properties, either the shortage or excess of certain elements results in soil problems.

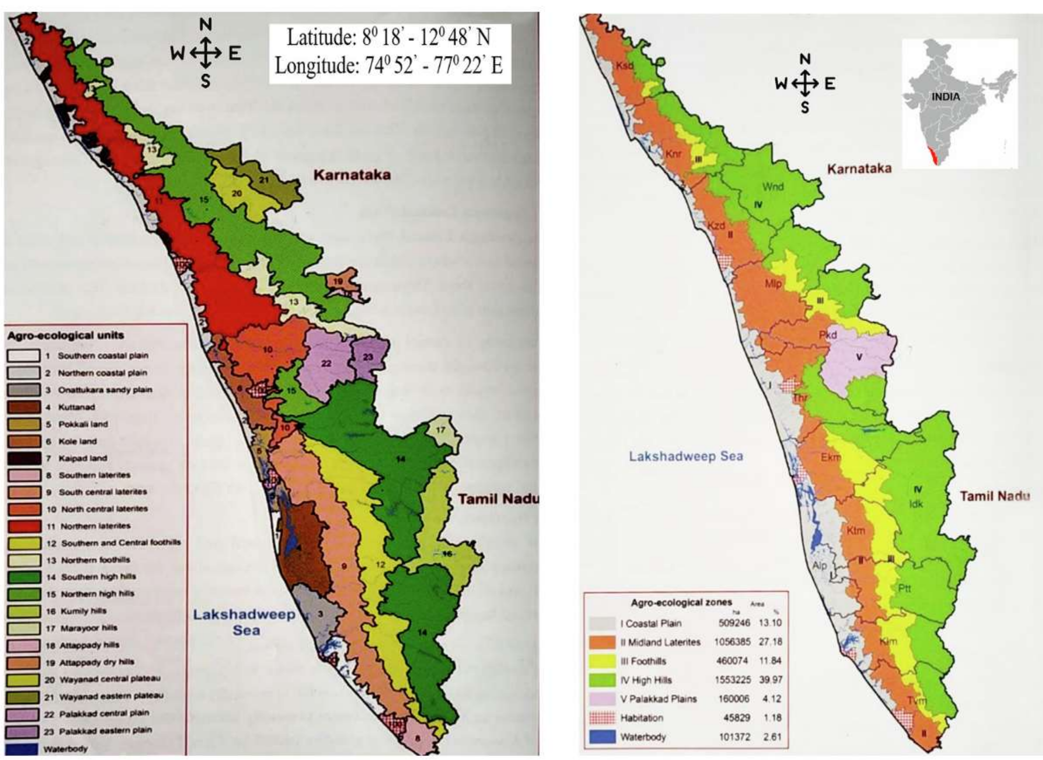

Fig. 1 – Soil map of Kerala [21,22]. The left figure is the division of AEZ and Midland Laterites is the selected zone of the current study and the right figure highlights the geographical study area North Central Laterites (AEU) denoted by the number 10.

Generally, these problems are extensive which limits agriculture production. It is necessary to place emphasis on AEU based approach for conservation and management of soil in a systematic method. It is recognized that integrating Information Technology with supporting services and inputs is important and will have a greater effect in addressing the extensive gaps in AEUs. The technological enhancement of the treatment of soil acidity, multiple nutrient deficiencies, and plant health management helps to revive the agriculture sector. This, in turn, produces an increase in productivity from the current levels. The village wise soil fertility indices help to maintain soil nutrients up to critical levels and these indices are used to know about the natural excess and deficiency in soil OC, P, K, and B.

Soils belonging to Kerala are naturally acidic due to intense leaching conditions. Estimate of plant available nitrogen in soils is often made by determining the organic carbon content of the oil. The State naturally contains a high level of organic matter. P is required by plants for energy transformation and photosynthesis. K plays a regulatory role in plant metabolism and development. In general, the available S is sufficient for the soils of Kerala. Zn influences translocation and transport of P in plants and it is required in very small quantities. Cu is a component of many essential enzymes and it is also required in small quantities. B plays a vital role in the physiology of the plants. The village wise soil fertility indices (FI) for the nutrients OC, P, K, and B is determined by the agriculture planning of the Indian Government using the threshold values to calculate their fertility levels as low, medium and high, which is listed in Table 1. Intervals for the major and micronutrients are defined as per the Package of Practices Recommendations [28].

Parker's nutrient index is used for the calculation of fertility index, which is used to associate soil conditions within a given area by classifying the region into various classes on the basis of a six-tier system [29]. The value of FI is the same for all the cultivation lands in any particular village.

$$\mathit{Soil\ Fertility\ Index}(FI) = \frac{(\mathrm{V.H}*3)+(\mathrm{H}*2.5)+(\mathrm{M.H}*2)+(\mathrm{M}*1.5)+(\mathrm{L}*1)+(\mathrm{V.L}*0.5)}{\text{Total number of cultivation lands}}$$

where, V.H, H, M.H, M, L, and V.L are denoted as the number of cultivation lands in the very high, high, moderately high, moderate, low and very low group in any particular village. The soil is categorized into these levels based on its chemical characteristics. Table 2 shows the six-tier ratings of soil OC and soil available primary nutrients P and K. Table 3 shows the six-tier ratings of soil available micronutrient Boron [30].

The village wise fertility index of OC, P, K, and B is indicated by the class labels as OC-F, P-F, K-F, and B-F. The values of indices are completely different from the input values of OC, P, K, and B. The village wise fertility index inputs used for our classification problems is listed in Table 4.

Table 1 – Threshold values for the village wise fertility indices [28].

Fertility index	Fertility level
<1.67	Low
1.67–2.33	Medium
>2.33	High

Table 3 – Six tier ratings of Boron. here ppm stands for parts per million [29].

Boron (ppm)	Fertility level ratings
<0.25	Very Low
0.25–0.5	Low
0.5–0.75	Medium
0.75–1.0	Moderately high
1.0–1.5	High
>1.5	Very high

Table 4 – Patterns per class used for each classification problem.

Classification problem	Classes	Number of cultivation lands per class
OC-F	Low	148
	Medium	309
	High	307
P-F	Low	680
	Medium	84
K-F	Low	513
	Medium	200
	High	51
B_F	Low	648
	Medium	116
pH	Strongly acidic	85
	Highly acidic	392
	Moderately acidic	214
	Slightly acidic	53

Table 2 – Six tier ratings of available primary nutrients (P, K) and OC (Organic Carbon) [29].

Organic Carbon (OC) (%)	Available soil nutrients (Kg ha^{-1})		Ratings
	Phosphorus (P)	Potassium (K)	
<0.20	<7	<100	Very Low
0.21–0.40	7.1–14	101–150	Low
0.41–0.60	14.1–21	151–200	Medium
0.61–0.80	21.1–28	201–250	Moderately high
0.81–1.0	28.1–35	251–300	High
>1.0	>35	>300	Very high

2.3. Classification of soil pH

Soils of the humid tropics are naturally acidic in reaction due to the intense leaching conditions and the consequent loss of basic cations. Strongly acid soils are a stressed environment for plant growth. The primary cause for the development of the strong acid condition in soils of Kerala is the heavy input of fertilizers without regular application of lime to neutralize the acidity generated. The patterns used for pH classification is listed in Table 4. Even though the ratings of pH in the soil are indicated in Table 5, for our pH classification problem it is sufficient to differentiate the 4 classes: Strongly Acidic (SA), Highly Acidic (HA), Moderately Acidic (MA), Slightly Acidic (SLA). To evaluate microbial processes, nutrient levels and to select appropriate crops and pesticides, the classification of pH into different ranges is very much essential [28,30].

Table 5 – Rating of pH by six-tier system [27].

pH (1:2.5)	Ratings
<4.5	Strongly acidic
4.6–5.5	Highly acidic
5.6–6.5	Moderately acidic
6.6–6.9	Slightly acidic
7.0	Neutral
7.1–8.0	Slightly alkaline
8.1–9.0	Moderately alkaline
9.1–10.0	Strongly alkaline
10.1–11.0	Very strongly alkaline

2.4. Proposed classification framework

The procedure described in Fig. 2 is used for implementing all the above classification models. The samples of each classification problem are arbitrarily rearranged and 80% of them are used for training and cross-validation and the remaining 20% for testing. Hence the tenfold cross-validation strategy is used here for training and validation, wherein each fold, 90% of training data is devoted for training and 10% of the same for validation. Each ELM classifier is trained on the training set using a different combination of parameters i.e., training function and the number of hidden nodes, and then it is verified on the validation sets. The best parameters are calculated and selected from the training set and are then used for testing the data. The final test result obtained is considered as the output of the corresponding classifier for the analysis, which is averaged over ten trials. For the ELM classifiers, one of the meta-parameter used for optimization is the number of hidden neurons. This parameter is fixed to fifty for soil nutrients classification by tuning the values in the range [10, 150] and fixed to one hundred and fifty for pH classification by tuning in the range [10, 200]. Another meta-parameter used for optimizing the ELM model is the activation function. Activation functions used for this study are the gaussian radial basis, sine-squared, hyperbolic tangent, triangular basis, and hard limit. The training data with each activation function is applied to the ELM model for classifying the data and the accuracy obtained with these models are compared. The activation function with maximum accuracy is selected as the optimal classifier meta-parameter. Using the tuned value

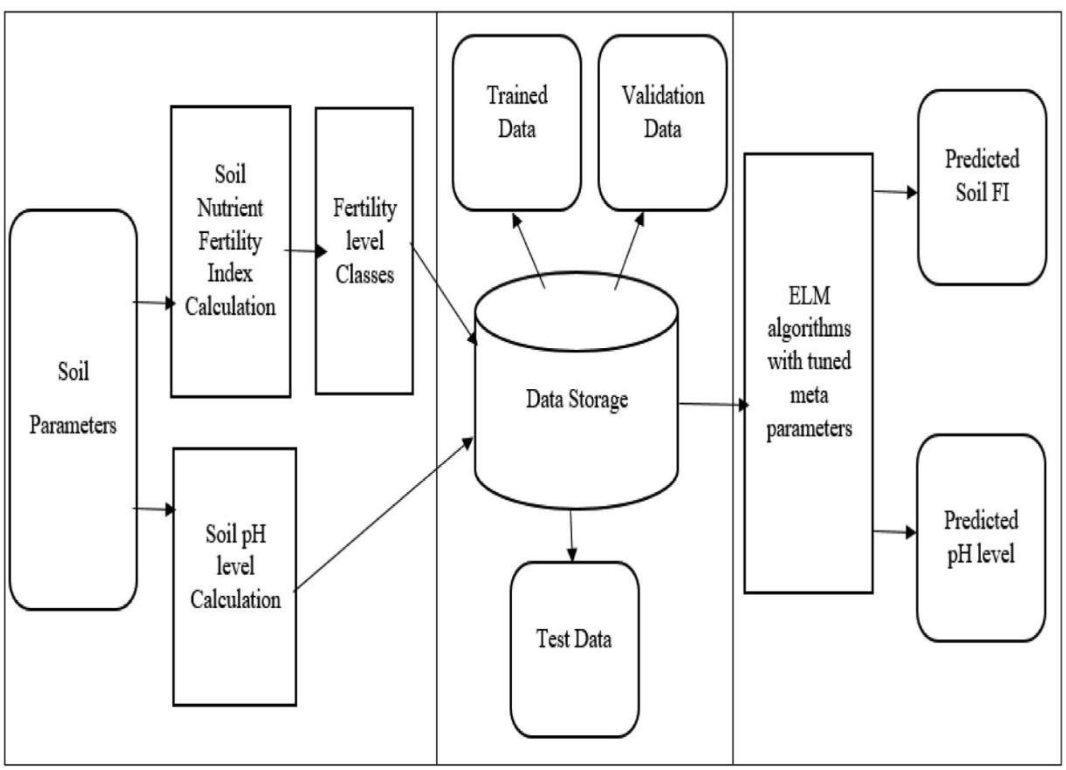

Fig. 2 – Proposed workflow for soil parameter classification and prediction.

of the number of hidden neurons and activation function, the soil nutrients and pH are classified and predicted with more accuracy.

2.5. Extreme learning machines

In the current study, the classifier used is selected based on the learning algorithm which can learn much quicker than commonly used learning algorithms i.e., Extreme Learning Machine [18]. The best generalization performance is achieved in this learning algorithm for feedforward neural networks. Even though this algorithm is not complicated, at the same time it is a highly efficient learning algorithm for single-hidden layer feedforward neural networks (SLFNs). It can choose the input weights randomly without any preconceived notion and analytically decides the output weights of SLFNs. This learning algorithm is executed by the publically available Python code which is based on scikit-learn [31]. According to popular observation, the ELM learning algorithm could be used to train SLFNs with several activation functions which are non-differentiable and non-regular [32]. The hyperbolic tangent function (elm_tanh), triangular basis transfer function (elm_tribas), hard limit transfer function (elm_hardlim), gaussian radial basis function (elm_grbf), sine-squared function (elm_sinsq) and many non-regular functions are some of the activation functions which can be noted [33]. Theoretically, SLFNs can implement any classification application and approximate any continuous function [34]. A unified learning platform is provided by ELM with a widespread type of feature mapping techniques. This can be applied in function approximation and multiclass classification application directly by optimizing the number of neurons and the activation function [35]. ELM tends to achieve the smallest norm of weights and the smallest training error [36]. In the study which is being undertaken, the above-mentioned soil classification problems are implemented by ELMs with different activation functions. Apart from that, the performance of each classification is compared with the other.

2.6. Experimental setup

The accuracy score which computes the accuracy by the count of correct predictions is the measure used for classification performance. The fraction of correct predictions over $k_{samples}$ is defined as

$$accuracy(Y, y) = \frac{1}{k_{samples}} \sum_{i=0}^{k_{samples}-1} 1(y_i = Y_i) \quad (1)$$

where y_i is the predicted value of the i^{th} sample, Y_i is the corresponding true value and 1(x) is the indicator function [37]. The other four measures used for calculating the classification and prediction performance are Kappa, Precision, Recall and F Score [38,39].

$$Kappa(K) = \frac{p_O - p_e}{1 - p_e} \quad (2)$$

where p_o is the probability of the actual observed value of the classification and p_e is the probability of a right classification by chance. The confusion matrix obtained after the methodology analysis helps to compute the True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) values.

$$Precision(\%) = \left(\frac{TP}{TP + FP}\right)100 \quad (3)$$

$$Recall(\%) = \left(\frac{TP}{TP + FN}\right)100 \quad (4)$$

$$FScore(\%) = \left(\frac{2(Precision * Recall)}{Precision + Recall}\right)100 \quad (5)$$

3. Results and discussion

3.1. Model building using North Central Laterite region datasets

The information about each classification task and the universal set of each class is listed in Table 4. North Central Laterites region has soil with K-F level in medium and low classes and OC-F level covers medium and high classes with equal probability. This soil contains low P-F and B-F levels; the low fertility level class is the most occupied for both classification problems. It is highly acidic, so the HA class is the most populated. North Kerala is highly prone to soil erosion, leaching of soil nutrients, floods and droughts, so predicting the type of soil pH can avoid the unnecessary usage of chemical fertilizers. The data are preprocessed in order to have one standard deviation and zero mean for each input. The cross-validated results of different classification problems are plotted in Fig. 3. The peak values of each graph in Fig. 3 indicate the best value that can be used as the number of hidden neurons for obtaining the maximum trained accuracy i.e., 50 is the optimal number of hidden neurons for soil nutrients classification and 150 for pH classification. The cross-validated accuracy score achieved by each classifier for 5 classification problems is illustrated in Fig. 3. From that, we can understand the best performance of soil nutrients classification occurred by the use of gaussian radial basis function and pH classification occurred by hyperbolic tangent function.

The accuracy score obtained for the testing data by using the optimized meta-parameter values is indicated in Table 6. Finally, elm_grbf achieved the best accuracy score for OC-F, P-F, K-F, and B-F classification problems, whereas elm_tanh is observed to be the best for pH classification problems.

The confusion matrices obtained during the testing period of data in all ELM classifiers, along with Accuracy score, Precision, Recall, and F Score are illustrated in Tables 7–11, for village wise soil fertility indices problems OC-F, P-F, K-F, and B-F.

The elm_grbf achieves the best accuracy score for OC-F (83.66), P-F (90), K-F (78.43) and B-F (88.23). The confusion matrices obtained during the testing period of data in all ELM classifiers, along with Accuracy score, Precision, Recall, and F Score are illustrated in Tables 12 and 13, for pH classification problem. The elm_tanh achieves the best accuracy score for pH (88.59).

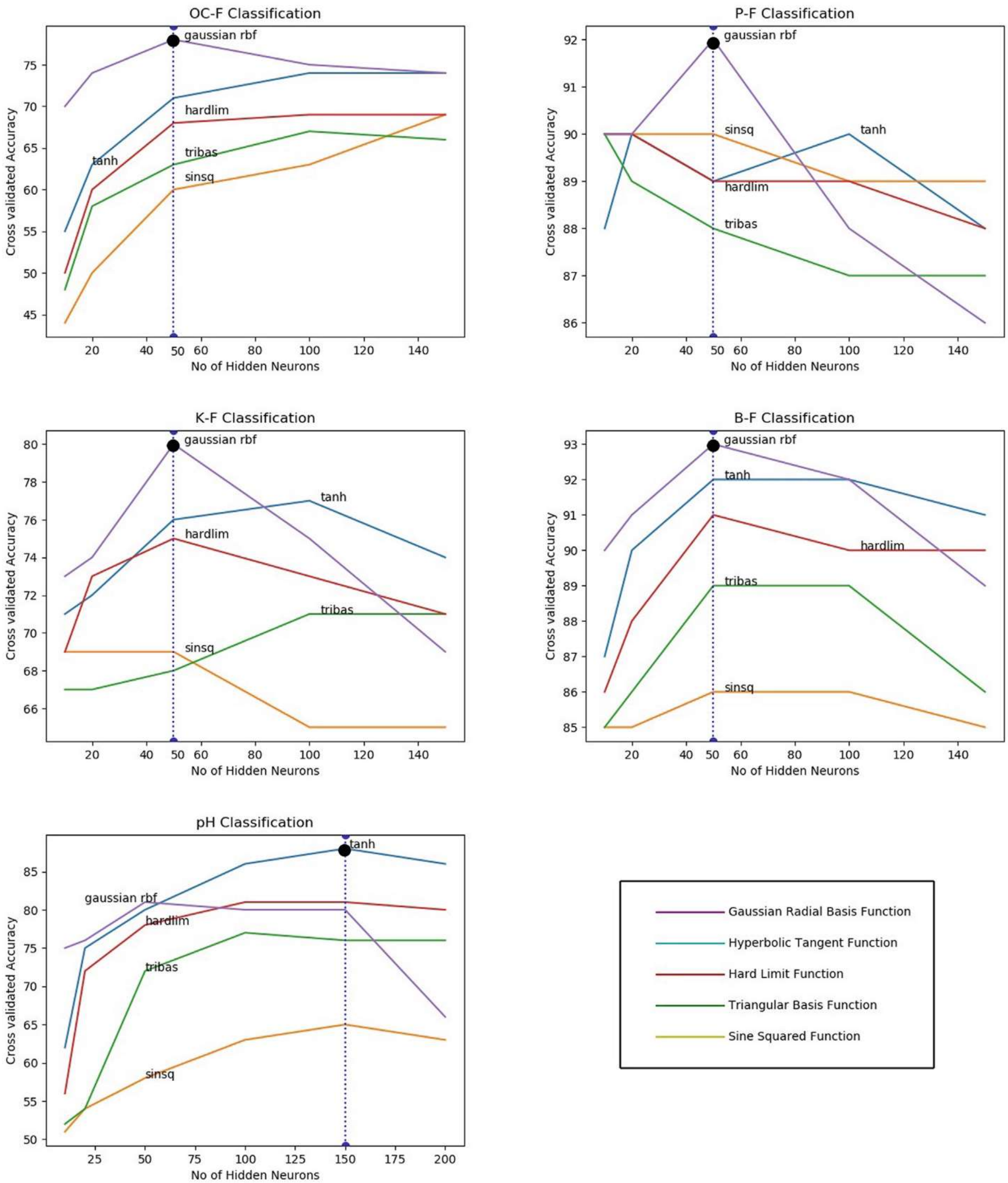

Fig. 3 – Graphical plotting of each cross-validated classification accuracy (in %) with a different number of hidden nodes and for different ELM activation functions. The dotted line is drawn at the peak which indicates the optimal number of hidden neurons.

3.2. *Model testing using Marathwada region datasets*

Soil fertility index classification and pH level classification of Marathwada Region of Maharashtra were analyzed by Sirsat MS et al. (2017) using ELM [15]. The outputs exhibit a moderate accuracy for ELM classifiers. To obtain the optimal accuracy value, ELM model was tuned with triangular basis activation function and 20 hidden neurons. When our proposed model is validated with this dataset, a better accuracy is obtained than the existing ELM model. The meta-parameters obtained after tuning the classification problems with North Central Laterite dataset are used for evaluating

Table 6 – Accuracy score rate (in %) obtained by each ELM classifier for each soil parameter classification problem and the best accuracy for each problem is given in bold type.

elm_classifier	Testing accuracy of soil nutrients				Testing accuracy of pH
	OC-F	P-F	K-F	B-F	
elm_tanh	81.04	88.23	71.89	86.27	**88.59**
elm_sinsq	67.32	86.92	62.74	84.96	71.14
elm_tribas	64.70	85.62	64.70	84.31	78.52
elm_hardlim	73.20	86.27	73.85	86.27	85.23
elm_grbf	**83.66**	**90.0**	**78.43**	**88.23**	81.87

Table 7 – Confusion matrix report of elm_tanh classifier for soil nutrients classification.

Fertility level	OC-F (elm_tanh)			P-F (elm_tanh)		K-F (elm_tanh)			B-F (elm_tanh)	
	Low	Medium	High	Low	Medium	Low	Medium	High	Low	Medium
Low	22	0	1	130	3	85	9	0	117	8
Medium	1	42	10	15	5	31	14	1	13	15
High	5	12	60	–	–	2	0	11	–	–
Accuracy score (%)		**81.04**		**88.23**			**71.89**		**86.27**	
Kappa (%)		69.26		30.52			40.74		50.68	
Precision (%)	79	78	85	90	62	72	61	92	90	65
Recall (%)	96	79	78	98	25	90	30	85	94	54
F-score (%)	86	79	81	94	36	80	41	88	92	59

Table 8 – Confusion matrix report of elm_sinsq classifier for soil nutrients classification.

Fertility level	OC-F (elm_sinsq)			P-F (elm_sinsq)		K-F (elm_sinsq)			B-F (elm_sinsq)	
	Low	Medium	High	Low	Medium	Low	Medium	High	Low	Medium
Low	22	0	1	133	0	87	6	1	123	2
Medium	6	27	20	20	0	39	7	0	21	7
High	7	16	54	–	–	10	1	2	–	–
Accuracy score (%)		**67.32**		**86.92**			**62.74**		**84.96**	
Kappa (%)		47.42		0			12.28		31.76	
Precision (%)	63	63	72	87	0	64	50	67	85	78
Recall (%)	96	51	70	100	0	93	15	15	98	25
F-score (%)	76	56	71	93	0	76	23	25	91	38

Table 9 – Confusion matrix report of elm_tribas classifier for soil nutrients classification.

Fertility level	OC-F (elm_tribas)			P-F (elm_tribas)		K-F (elm_tribas)			B-F (elm_tribas)	
	Low	Medium	High	Low	Medium	Low	Medium	High	Low	Medium
Low	23	0	0	131	2	86	7	1	119	6
Medium	11	23	19	20	0	38	8	0	18	10
High	2	22	53	–	–	5	3	5	–	–
Accuracy score (%)		**64.70**		**85.62**			**64.70**		**84.31**	
Kappa (%)		43.61		<0			20.38		37.08	
Precision (%)	64	51	74	87	0	67	44	83	87	62
Recall (%)	100	43	69	98	0	91	17	38	95	36
F-score (%)	78	47	71	92	0	77	25	53	91	45

the Marathwada dataset. In this study, ELM with gaussian radial basis activation function is found to be working effectively with 50 hidden neurons for soil nutrients classification and 150 hidden neurons for pH classification. Kappa statistics of Marathwada dataset obtained in all proposed ELM classifiers is compared with the existing classifier (Triangular basis transfer function), which is illustrated in Table 14. The elm_grbf achieves the best accuracy score for OC-F (77.66), P-F

Table 10 – Confusion matrix report of elm_hardlim classifier for soil nutrients classification.

Fertility level	OC-F (elm_hardlim)			P-F (elm_hardlim)		K-F (elm_hardlim)			B-F (elm_hardlim)	
	Low	Medium	High	Low	Medium	Low	Medium	High	Low	Medium
Low	20	0	3	132	1	87	7	0	117	8
Medium	2	34	17	20	0	30	16	0	13	15
High	4	15	58	–	–	3	0	10	–	–
Accuracy score (%)		**73.20**		**86.27**			**73.85**		**86.27**	
Kappa (%)		55.84		<0			44.06		50.68	
Precision (%)	77	69	74	87	0	72	70	100	90	65
Recall (%)	87	64	75	99	0	93	35	77	94	54
F-score (%)	82	67	75	93	0	81	46	87	92	59

Table 11 – Confusion matrix report of elm_grbf classifier for soil nutrients classification.

Fertility level	OC-F (elm_grbf)			P-F (elm_grbf)		K-F (elm_grbf)			B-F (elm_grbf)	
	Low	Medium	High	Low	Medium	Low	Medium	High	Low	Medium
Low	23	0	0	128	5	88	6	0	117	8
Medium	3	41	9	11	9	25	20	1	10	18
High	5	8	64	–	–	1	0	12	–	–
Accuracy score (%)		**83.66**		**90.0**			**78.43**		**88.23**	
Kappa (%)		73.58		47.26			55.42		59.53	
Precision (%)	74	84	88	92	64	77	77	92	92	69
Recall (%)	100	77	83	96	45	94	43	92	94	64
F-score (%)	85	80	85	94	53	85	56	92	93	67

Table 12 – Confusion matrix report of elm_grbf, elm_sinsq and elm_tribas classifier for pH level classification.

pH level	pH (elm_grbf)				pH (elm_sinsq)				pH (elm_tribas)			
	SA	HA	MA	SLA	SA	HA	MA	SLA	SA	HA	MA	SLA
SA	12	2	0	0	8	2	3	1	10	4	0	0
HA	5	69	8	1	2	69	9	3	3	75	5	0
MA	2	3	34	1	2	13	25	0	1	13	26	0
SLA	1	1	3	7	1	1	6	4	1	0	5	6
Accuracy score (%)			**81.87**			**71.14**				**78.56**		
Kappa (%)			70.82			51.27				62.87		
Precision (%)	60	92	76	78	62	81	58	50	67	82	72	100
Recall (%)	86	83	85	58	57	83	62	33	71	90	65	50
F-score (%)	71	87	80	67	59	82	60	40	69	86	68	67

Table 13 – Confusion matrix report of elm_hardlim and elm_tanh classifier for pH level classification.

pH level	pH (elm_hardlim)				pH (elm_tanh)			
	SA	HA	MA	SLA	SA	HA	MA	SLA
SA	10	4	0	0	12	2	0	0
HA	4	73	6	0	3	77	3	0
MA	1	3	36	0	0	5	34	1
SLA	0	1	3	8	0	0	3	9
Accuracy score (%)		**85.23**				**88.59**		
Kappa (%)		75.48				80.95		
Precision (%)	67	90	80	100	80	92	85	90
Recall (%)	71	88	90	67	86	93	85	75
F-score (%)	69	89	85	80	83	92	85	82

Table 14 – comparing the Kappa values (in %) obtained by each ELM classifier for each soil parameter classification problem in Marathwada dataset [15] with our proposed ELM classifiers.

elm_classifier	Activation function	Testing accuracy of soil nutrients				Testing accuracy of pH
		OC-F	P-F	Mn-F	Fe-F	
Marathwada data set [15]	Triangular basis transfer function	75.45	68.99	48.41	49.90	34.61
Marathwada data set with proposed ELM classifiers	Gaussian radial basis activation function	**77.66**	**69.98**	**57.10**	**65.14**	73.30
	Sine-squared function	12.69	6.42	0	4.87	45.45
	Hyperbolic tangent function	73.00	57.17	48.50	56.25	**78.99**
	Triangular basis function	28.87	26.74	19.64	15.73	60.88
	Hard limit function	64.60	41.78	46.88	48.13	70.71

(69.98), Mn-F (57.10) and Fe-F (65.14). The elm_tanh achieves the best accuracy score for pH (78.99).

4. Conclusions

The main reason for a heavy loss in soil quality is due to the incorrect soil and crop management strategies. Also, excess use of chemical fertilizers has created imbalances in the availability of soil nutrients. The productivity of Kerala soils is affected by all these factors. It is essential to place emphasis on AEU based approach for management and conservation of natural resources like soil in a systematic method. Soil resource management is an issue to be addressed at the AEU level. Widespread deficiency of secondary and micronutrients is observed in several Kerala AEUs. To identify the constraints that have resulted in the production gaps, the region-wise researches are required for any crop and soil. North Kerala is highly prone to soil erosion, leaching of soil nutrients, floods, and droughts. The main objective of this work is to classify area wise soil fertility indices based on the village level soil fertility information, which can be used in making a village wise soil fertility index analysis report. The current study aids the Kerala Government to make decisions about improving crop production and soil quality. The soil quality depends on its pH, EC, primary, micro and macronutrients on the selected crop. Hence, the classification of the pH and relevant soil nutrient indices help to save the time of specialized technicians and analyze soil health and environmental quality. It is recognized that integrating information with supporting services will have a greater effect in addressing the extensive gaps in AEUs. Machine Learning techniques help in the domain of agriculture and could support data analysis for making predictions. This study analyses the soil problems in Kerala region and interprets the soil test results in an efficient manner.

The very fast and simple learning algorithm called ELM classifier with its different activation functions are used to predict the fertility levels as low, medium and high using the soil parameters as input values; and to predict the pH levels of the corresponding soil. We analyze the prediction accuracy of different ELM classifiers with the activation functions which include the hyperbolic tangent function (elm_tanh), triangular basis transfer function (elm_tribas), hard limit transfer function (elm_hardlim), gaussian radial basis function (elm_grbf) and sine-squared function (elm_sinsq). We achieved accuracy score rate above 90% for the village wise P fertility index level and above 85% for the micronutrient B fertility index level, while the classification accuracy of the village wise OC fertility index level achieved above 80%. The accuracy score rate is about 78% for K fertility index. The best accuracy value obtained for pH classification is about 89%. Finally, the ELM with gaussian radial basis activation function achieves the best performance for soil nutrient fertility index classification, both in terms of accuracy rate and Kappa, followed by hyperbolic tangent, hard limit, triangular basis, and sine-squared functions. The ELM with hyperbolic tangent function achieves a good result for pH classification followed by gaussian radial basis, hard limit, and triangular basis functions. The python-version of ELM with gaussian radial basis function provides a better result in the classification of the soil problems, as it is the best in four of five problems and which goes above 80% in most of the accuracy rate calculations in all the cases. We also examined the classification model accuracy across Marathwada region, Maharashtra, India by training and testing each classifier with data from this region. It is then validated with the dataset and the optimized meta-parameters for classification has effectively worked in this case also.

The overall objective of this study is to develop a neural network model to classify and predict soil fertility indices and pH values. The outcomes of this work might help to make a machine learning decision system for Kerala Government to manage the soil nutrient deficiency problems. Results showed that optimization of ELM parameters helps to create a suitable model for soil fertility index classification. In the future, the classification of soil nutrients N_2O, P_2O_5, and K_2O using efficient and fast ELM methods or any advanced neural network techniques can be used for fertilizer recommendation for the required crop. These methods can also be used to consider fertility indices of other nutrients available and to create soil fertility maps. This proposed model can be applied in another agro-ecological region to analyze the soil parameters.

Declaration of Competing Interest

The authors declare no conflict of interest.

REFERENCES

[1] Sheela PJ, Sivaranjani K, Phil M. A brief survey of classification techniques applied to soil fertility prediction. Int Conf Eng Trends Sci Hum 2015:80–3.

[2] de Paul Obade V, Lal R. Towards a standard technique for soil quality assessment. Geoderma 2016;265:96–102.

[3] Mucherino A, Papajorgji P, Pardalos PM. A survey of data mining techniques applied to agriculture. Oper Res Int J 2009;9(2):121–40.

[4] Romero JR, Roncallo PF, Akkiraju PC, Ponzoni I, Echenique VC, Carballido JA. Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires. Comput Electron Agric 2013;96:173–9.

[5] Pantazi XE, Moshou D, Alexandridis T, Whetton RL, Mouazen AM. Wheat yield prediction using machine learning and advanced sensing techniques. Comput Electron Agric 2016;121:57–65.

[6] Hill MG, Connolly PG, Reutemann P, Fletcher D. The use of data mining to assist crop protection decisions on kiwifruit in New Zealand. Comput Electron Agric 2014;108:250–7.

[7] Ritz C, Putku E, Astover A. A practical two-step approach for mixed model-based kriging, with an application to the prediction of soil organic carbon concentration. Eur J Soil Sci 2015;66(3):548–54.

[8] Schillaci C, Acutis M, Lombardo L, Lipani A, Fantappie M, Märker M, et al. Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: the role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. Sci Total Environ 2017;601:821–32.

[9] Wang B, Waters C, Orgill S, Cowie A, Clark A, Li Liu D, et al. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. Ecol Ind 2018;88:425–38.

[10] Terhoeven-Urselmans T, Vagen TG, Spaargaren O, Shepherd KD. Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. Soil Sci Soc Am J 2010;74(5):1792–9.

[11] Jia HY, Chen J, Yu HL, Liu DY. Soil fertility grading with Bayesian Network transfer learning. In: International conference on machine learning and cybernetics. IEEE; 2010. p. 1159–63.

[12] Veronesi F, Korfiati A, Buffat R, Raubal M. In: Assessing accuracy and geographical transferability of machine learning algorithms for wind speed modelling. Cham: Information Science. Springer; 2017. p. 297–310.

[13] Elavarasan D, Vincent DR, Sharma V, Zomaya AY, Srinivasan K. Forecasting yield by integrating agrarian factors and machine learning models: A survey. Comput Electron Agric 2018;155:257–82.

[14] Reashma SJ, Pillai AS. Edaphic factors and crop growth using machine learning—A review. In: International conference on intelligent sustainable systems (ICISS). IEEE; 2017. p. 270–4.

[15] Sirsat MS, Cernadas E, Fernández-Delgado M, Khan R. Classification of agricultural soil parameters in India. Comput Electron Agric 2017;135:269–79.

[16] Sirsat MS, Cernadas E, Fernández-Delgado M, Barro S. Automatic prediction of village-wise soil fertility for several nutrients in India using a wide range of regression methods. Comput Electron Agric 2018;154:120–33.

[17] Krishnan P, Venugopal KR, Sehgal J. Soils of Kerala for optimising land use. NBSS Publication; 1996. p. 48.

[18] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing 2006;70(1–3):489–501.

[19] Bhalla GS, Singh G. Economic liberalisation and Indian agriculture: a statewise analysis. Econ Polit Weekly 2009;26:34–44.

[20] Indian Government. Soil Health Card data. https://soilhealth.dac.gov.in/. 2017.

[21] Sehgal JL, Mandal DK, Mandal C, Vadivelu S. Agro-ecological regions of India. NBSS Publication; 1990. p. 24.

[22] Soil Fertility Assessment and Information Management for Enhancing Crop Productivity. In: Rajasekharan P, Nair KM, Rajasree G, Sureshkumar P, Narayanankutty MC, editors. Agro-ecology of Kerala. Kerala State Planning Board; 2013. p. 54–71.

[23] Kerala Agricultural University. Package of Practices Recommendations. http://pop.kau.in/appendix14.htm. 2011.

[24] Walkley A, Black IA. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Sci 1934;37(1):29–38.

[25] Bray RH, Kurtz LT. Determination of total, organic, and available forms of phosphorus in soils. Soil Sci 1945;59 (1):39–46.

[26] Gupta UC. A simplified method for determining hot-water soluble boron in podzol soils. Soil Sci 1967;103(6):424–8.

[27] Suchithra MS, Pai ML. Improving the performance of sigmoid kernels in multiclass SVM using optimization techniques for agricultural fertilizer recommendation system. In: International Conference on Soft Computing Systems. Singapore: Springer; 2018. p. 857–68.

[28] Kerala Agricultural University. Package of practices recommendations: crops. 14th ed. Thrissur; 2011.

[29] Ramamoorthy B, Bajaj JC. Available nitrogen, phosphorus and potassium status of Indian soils. Fertil News 1969.

[30] Indragir GN. GPS-GIS Based Soil Fertility Maps of Shirpur Tehsil of Dhule District (MS). Doctor thesis. Rahuri: Mahatma Phule Krishi Vidyapeeth. 2015.

[31] Extreme Learning Machines. http://www.extreme-learning-machines.org.

[32] Huang GB, Zhu QY, Mao KZ, Siew CK, Saratchandran P, Sundararajan N. Can threshold networks be trained directly? IEEE Trans Circuits Syst II Express Briefs 2006;53(3):187–91.

[33] Huang GB, Babri HA. Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. IEEE Trans Neural Networks 1998;9(1):224–9.

[34] Huang GB, Chen YQ, Babri HA. Classification ability of single hidden layer feedforward neural networks. IEEE Trans Neural Networks 2000;11(3):799–801.

[35] Huang GB, Zhou H, Ding X, Zhang R. Extreme learning machine for regression and multiclass classification. IEEE Trans Syst, Man, Cybernet, Part B (Cybernetics) 2012;42 (2):513–29.

[36] Fernández-Navarro F, Hervás-Martínez C, Sanchez-Monedero J, Gutiérrez PA. MELM-GRBF: a modified version of the extreme learning machine for generalized radial basis function neural networks. Neurocomputing 2011;74 (16):2502–10.

[37] Sokolova M, Japkowicz N, Szpakowicz S. Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In: Australasian Joint Conference on Artificial Intelligence. Berlin, Heidelberg: Springer; 2006. p. 1015–21.

[38] Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005;37(5):360–3.

[39] Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. J Mach Learn Technol 2011;2(1):37–63.