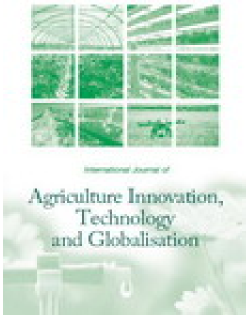
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Agriculture Innovation, Technology and Globalisation

ISSN online: 2516-1970 - ISSN print: 2516-1962
https://www.inderscience.com/ijaitg

Extreme learning machine for solving paddy nutrient deficiencies in Davangere region

M. Varsha, M.P. Pavan Kumar, S. Basavarajappa

DOI: 10.1504/IJAITG.2023.10057454

Article History:

Received:	02 September 2022
Last revised:	26 February 2023
Accepted:	17 March 2023
Published online:	28 November 2023

Extreme learning machine for solving paddy nutrient deficiencies in Davangere region

M. Varsha*

Bapuji Institute of Engineering and Technology,
Davangere, India
Email: mvarsha16@gmail.com
*Corresponding author

M.P. Pavan Kumar

Jawaharlal Nehru National College of Engineering,
Shimoga, India
Email: pavankumarjnnce@gmail.com

S. Basavarajappa

G.M. Institute of Technology,
Davangere, India
Email: basuinbetur@gmail.com

Abstract: Soil nutrient is an important aspect that contributes to the soil fertility and environmental effects. Traditional evaluation approaches of soil nutrient are quite hard to operate and they are very slow, making great difficulties in practical applications. The proposed study, presents extreme learning machine (ELM) for analysing the soil fertility index values of boron, zinc, organic carbon and pH in Davangere District. Boron, zinc, organic carbon, and pH concentrations in soil play significant roles in paddy crop cultivation and growth. Proposed ELM-based approach helps in the prediction of boron, zinc, organic carbon and pH index values in soil by evaluating four linear and nonlinear activations functions. Performance of ELM model is analysed by increasing the number of hidden neurons in the hidden layer.

Keywords: extreme learning machine; ELM; transfer functions; hidden neurons; back-propagation.

Reference to this paper should be made as follows: Varsha, M., Kumar, M.P.P. and Basavarajappa, S. (2023) 'Extreme learning machine for solving paddy nutrient deficiencies in Davangere region', *Int. J. Agriculture Innovation, Technology and Globalisation*, Vol. 3, No. 2, pp.157–176.

Biographical notes: M. Varsha received her BE in Computer Science and Engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India, MTech in Computer Networks from Visvesvaraya Technological University, Belagavi, Karnataka, India in 2013 and PhD in Computer Science and Engineering from Visvesvaraya Technological University, Belagavi in 2022, Karnataka. She is currently working as an Associate Professor, Department of Information Science and Engineering at Bapuji Institute of Engineering and Technology Davangere, India. Her area of research includes computer vision, machine learning, deep learning. She had published ten articles in various reputed journals' (Springer, and IGI Global). She also served as a reviewer for many conferences.

M.P. Pavan Kumar received his BE in Information Science and Engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India, MSc in Engineering by Research in Computer Science from University of Mysore, Mysuru in 2013, Karnataka, India and PhD in Computer Science and Engineering from Visvesvaraya Technological University, Belagavi in 2021, Karnataka. He is currently working as an Associate Professor, Department of Information Science and Engineering at Jawaharlal Nehru National College of Engineering, Shivamogga, India. His area of research includes computer vision, machine learning, human behaviour analysis and human centric computing. He had published 15 articles in various reputed journals' (Springer, World Scientific, Indescience, and IGI Global). He is also serving as a reviewer for MTAP, Springer and Journal of Supercomputing Springer.

S. Basavarajappa received her BE in Mechanical Engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India and MTech in Thermal Power Engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India in 2013. Currently, she is pursuing her PhD in Mechanical Engineering from Visvesvaraya Technological University, Belagavi in 2023, Karnataka. He is currently working as an Assistant Professor, Department of Mechanical Engineering G.M. Institute of Technology Davangere, India. He had published five articles in various reputed journals' (Scopus Indexed). He also served as a reviewer for many journals.

1 Introduction

The production of rice crop is affected by many factors which includes disease, pests and soil nutrients deficiency. Soil nutrients are very essential throughout the growth of paddy plant. Important soil nutrients of paddy crop are nitrogen, potassium, zinc, boron, organic carbon, etc. Nutrient's deficiency affects growth of the plant, in turn produces yellow leaves, reduces height of the plant and reduces yield. To avoid decline of soil nutrients, famers use excess of fertilisation that leads to the occurrence of diseases.

The management as well as enhancement of dynamic soil properties is the main emphasis of agro-based soil management for increasing paddy crop production. In developing country like India, rapid urbanisation, land constraints, and the demise of conventional soil farming practices have resulted in decline of fertility level of soil. Increased paddy production could be achieved by better soil resource management and micronutrient application optimisations. Decision-makers and small farmers choose suitable agricultural eco system management and soil resource development whenever problems with crop yield markers are detected and controlled in a timely manner. Machine learning (ML) approaches can now effectively handle prediction and classification challenges. The use of ML technologies in agriculture has significantly reduced the problems faced by farmers.

Karnataka state is well-known for the paddy cultivation, especially Mandya, Bellary Davangere, and Shimoga districts. In these districts varieties of seeds are cultivated in two season of year. The Monsoon season or Kharif Season (July–December) and Summer or Rabi Season (January–June). Every year, it has been reported that paddy fields in Davangere district are decline due to diseases or pests. Davangere district has mixed red and clay soil which contains less organic carbon, boron and zinc nutrients. Plants with boron and zinc deficiency are more susceptible to diseases and reduces yield. Excess of fertilisation increases the cost of paddy production and farmers cannot able to receive expected yield. Therefore, it is necessary to know levels of organic carbon, pH, boron and zinc nutrients.

Many previous works have considered image recognition and ML models to classify the images into healthy and unhealthy images. However, most of these algorithms require image segmentation and feature extraction. But, from the many extracted features, it is difficult to judge the important and dominant features for plant disease detection. Moreover, under difficult background circumstances, many techniques fail to successfully segment the leaf and will lead to unreliable deficiency recognition. So, image segmentation and feature extraction are still challenging tasks. Therefore, automatic plant disease detection and nutrient deficiency recognition are still challenging tasks. Recently, ML techniques is becoming the preferred scheme to overcome few challenges.

2 Related work

In recent advancements, through the application of ML many innovative and incredible works have been reported on disease detection and quantification of crops. de Paul Obade and Lal (2016) have proposed artificial neural networks (ANNs) which is called as back-propagation neural networks to estimate fertility level of soil when ML techniques was invented. In article, Romero et al. (2013) have used attributes such as availability of water content in the soil, electrical conductivity (EC), type of soil, and soil density. Partial least squares regression is developed to predict the fertility level of the soil. Many researchers and scientists have developed ML approaches and have proposed solutions for soil management in the field of agriculture, such as predicting soil fertility, supplying appropriate nutrient levels and water, and so on. Using phenotypic plant variables as inputs, Pantazi et al. (2016) used J48, K-nearest neighbours (KNN), One-R, and apriori classifier algorithms to classify and forecast wheat yield. Yield of the wheat was classified into three different class such as high, medium and low, considering these classes in target variable conventional counter propagation neural networks which is a supervised technique was used for classification by Hill et al. (2014). To make recommendations regarding insecticide treatment for kiwifruit, Ritz et al. (2015) used the tree based, regression based supervised techniques along with SVM and boosting algorithm. The soil organic carbon was predicted using an unbiased linear predictor in Schillaci et al. (2017). Boosted regression trees (Wang et al., 2018) are proposed by the researcher to predict the amount of OC in Sicilian soils. To estimate content of organic carbon in soils, the random forest was integrated with the genetic algorithms for feature selection (Terhoeven-Urselmans et al., 2010) Partial least squares were used to predict content of organic carbon, pH level in the different types of soil (Jia et al., 2010).

Most of the researchers have put efforts to assess the fertility level of each type of soil based on acidity content in the soil and other nutrients level in the soil (Mucherino et al., 2009). To raise and decrease the wind speed in the field, Veronesi et al. (2017) have implemented ML methods. Growth of plant in precision farming is directly dependent on soil fertility and climate change. Effects of these on different crops were analysed by Elavarasan et al. (2018) using tree based, neural network classifiers. ML techniques have been proposed (Reashma and Pillai, 2017) for the prediction of type of soil and level of soil nutrients. Sirsat et al. (2017) employed regression analysis pseudo transfer function to predict village wise fertility indices. Nowadays, farmers are using soil fertility data to make judgements on fertiliser quantity, fertiliser distribution procedures and fertiliser management for different types of soil.

Many researchers are investigating the efficiency of various algorithms in classifying and prediction of soil fertility levels. Sirsat et al. (2018) have demonstrated neural network model by training various activation functions in terms of dynamic, continuous and constrained results shows that continuous activation function in neural network can handle complex data. Krishnan et al. (1996) improved results of Sirsat et al. (2018) by using non-polynomial activation function in feed forward networks which estimates continuous functions. Due to slow convergence of backpropagation based neural networks for solving real world problems, Huang et al. (2006) investigated efficiency of SLFN neural network with n hidden nodes and nonlinear activation functions for learning finite data. Huang et al. (2012) defined working procedure and theoretical observations of SLFN with *n* hidden nodes by randomly assigning weights and bias for learning *N* inputs. Simulation results have proven that SLFN neural networks speeds up trained model and achieves better generalisation.

Many advanced and complex algorithmic models have resulted significant advances in ML (Singh et al., 2018; Kodaty and Halavath, 2021; Sethy et al., 2020; Islam et al., 2020; Jose et al., 2021). These extensively proposed algorithms can be used to predict soil nutrients deficiency in the specific region. The key objective of the proposed study is to predict soil nutrients level based on the region's soil fertility information and it would be used as fertilisers recommendation with the decision support system.

The proposed study is as follows:

1 A benchmark soil nutrients dataset is obtained for Davangere district which contains 17,614 instances with 16 attributes.

2 Extreme learning machine (ELM) architecture is used for classification and performance of ELM is evaluated for each soil nutrient.

3 The activation functions are evaluated to find best activation function for ELM architecture.

3 Methodology

The proposed ML architecture helps farmers to take right decision in terms of application of fertilisation during the period of paddy cultivation. The 80% of samples in the dataset are used for training and validation and remaining 20% is used to test ELM architecture. ELM is trained using different transfer or activation functions and number of hidden neurons. The first hyper-parameter used to optimise the ELM model is number of hidden neurons. Hyper-parameter number of neurons is varied in the range 20 to 300. Another hyper-parameter is transfer function which is used to optimise ELM model. Four linear and nonlinear transfer functions are used for predicting soil nutrients level and results of each are compared.

3.1 Soil parameters

Present work studies the geographical region located at North latitudes of 14,013’11.0” and 14,033’52.3” and the east longitudes of 75,048’53.9” and 76,009’28.3”, covering area around 5,975 square kilometres. The shift in a paddy agriculture economy indicates that paddy cultivation in the region of Karnataka has undergone significant structural changes. Paddy cultivation however, continues to be a key source of revenue for the people of Davangere region. The soil testing laboratory of Taralabalu Krishi Vigyana Kendra (TKVK) collects soil samples from individual farmers. Sample information in terms of soil attributes is made available to public. The soil samples collected for the proposed study is comprised of 16 parameters that are immediately relevant to plant soil nutrition: pH, electric-conductivity, organic-carbon, primary nutrients of crop growth (N, P, K), secondary nutrients of crop growth, and micronutrients such boron, zinc, etc. In the proposed study, soil samples test report benchmark dataset is obtained from soil health card dashboard which is comprised of 17,614 instances with 16 soil attributes.

Figure 1 Methodology of the proposed work (see online version for colours)

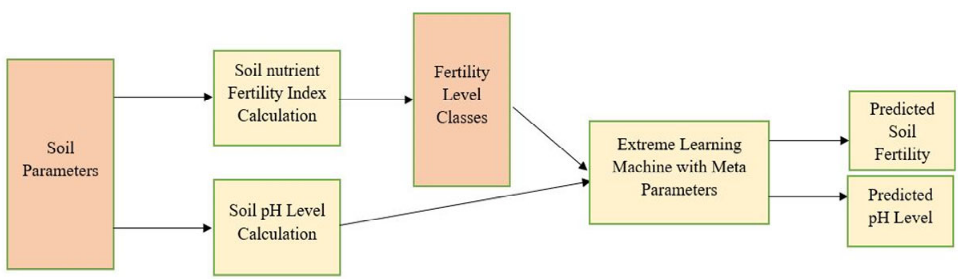


Figure 2 Overview of the study area (see online version for colours)

3.2 Soil fertility index values calculation and classification

Inappropriate soil-crop management practices during the growth cycle of plant have resulted in significant soil quality degradation. Excessive usage of synthetic fertilisers has thrown off the plant nutrients existence. Recent days, farmers are not giving attention towards assessing acidity content in the soil. Production of rice crop is threatened by a variety of factors, including a lack of micronutrients, primary nutrients and inefficient practices adapted by farmers for disease management. Most of the field crops such as wheat and rice suffer from micro-nutrients deficiency such as boron and zinc because field crops are frequently grown on soils with high pH alkaline soils and less organic carbon. Two other short comes excessively in paddy fields area are lack of cost-effective boron and zinc enriched nutrients fertiliser and unwillingness in farmers to fertilise rice fields with boron and zinc exacerbates

Table 1 Five tier rating of organic carbon

Fertility index	*Fertility level*
Very low	≤0.20
Low	0.21-0.59
Moderate	0.60-0.79
High	0.80-1.00
Very high	>1.00

Table 2 Two tier rating of boron nutrient

Fertility index	*Fertility level*
Sufficient	>0.50
Deficient	≤0.50

Table 3 Two tier rating of zinc nutrient

Fertility index	*Fertility level*
Sufficient	>0.50
Deficient	≤0.50

Table 4 Nine tier rating of pH

Fertility index	*Fertility level*
Strongly acidic	<4.5
Highly acidic	4.6-5.5
Moderately acidic	5.6-6.5
Slightly acidic	6.6-6.9
Neutral	7.0
Slightly alkaline	7.1-8.0
Moderately alkaline	8.1-9.0
Strongly alkaline	9.1-10.0
Very strongly alkaline	10.1-11.0

In general, the above-mentioned issues are widespread and limiting agricultural production. It is critical to emphasise the need of a ML based approaches to solve problems of soil fertility management and sustainability in a scientific manner. Southern transitions zone (VII) is known for its infertility level in soil. It is widely acknowledged that integrating information technology with supporting services plays critical role, and doing so, will have a bigger impact on closing STZ (VII) infertility. Estimation of fertility level of soil nutrients in Davangere district helps intensively in increasing paddy production. The soil fertility indices (FI) for the micronutrients Zn and B are defined by two levels deficient and sufficient, organic carbon are defined by 5 levels and pH is defined by 9 levels. Nutrients levels are shown in the Tables 1 to 4.

Parker's nutritional index is being used to calculate the fertility index that can be used to correlate soil characteristics within a given location by grouping the region into certain classes. The fertility index value is same across cultivation lands in any specified region. Soil fertility index is defined in equation (1)

$$Soil\ \ fertility\ index\ (FI) = \frac{\begin{array}{c}(VH \times 3) + (H \times 2.5) + (MH \times 2) \\ +(M \times 1.5) + (L \times 1) + (VL * 0.5)\end{array}}{Total\ number\ of\ \ cultivation\ lands} \tag{1}$$

where *VH*, *H*, *MH*, *L*, *M* and *VL* are defined as cultivation lands in the very high, high, moderately high, moderate, low and very low group of any districts.

3.3 Extreme learning machine

ELM is based on the concept of training only one layer of feed forward networks. Randomised strategy is being used towards assignment of weights and bias values between input and hidden layers. The ELM model uses nonlinear activation functions in the hidden layer to make system as nonlinear. Weights and bias values between hidden and output layer are learnt during model training which makes system to converge faster than conventional approaches. Conceptual analysis reveals that ELM has capability to attain the optimum solution with random variables than conventional networks with all characteristics to be learned. Other than above characteristics of ELM has ability for the estimation of complex translation from input values to nonlinear values and handles the large classes which are difficult in case of conventional models.

Figure 3 Basic structure of single layer feed forward network (see online version for colours)

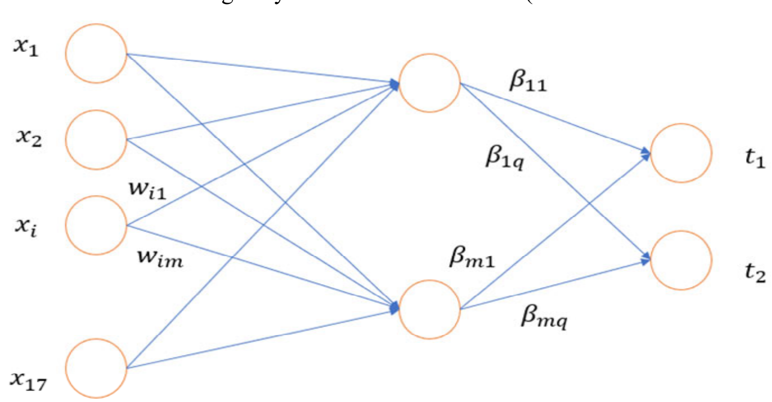


Table 5 Model parameters proposed in the study

Model parameter	*Parameter quality*
Input shape	2 dimensional
Output	Binary and multi-class classification
Neurons	20–300
Weights	Random
Biases	Random
Activation functions	Linear and nonlinear

The proposed study aims in classifying and prediction of secondary, micro-nutrients and pH levels using ELM. To achieve this, SLFN model is trained by random weights and bias values in the hidden layer with four linear and nonlinear activation functions. Even though SLFNs is called as nonlinear system after assigning random weights and bias values ELM model can simply considered as a linear system. Figure 3 gives architecture of single layer feed forward network. Weights of the output layer is determined by Moore Porse inverse generalisation function. Based on the above concepts, the ELM is used for the prediction of paddy crop deficiency levels, a simplest SLFNs model which learns thousands of times faster than conventional neural network algorithm in achieving better generalisation performance. To distinguish SLFN from other learning algorithms, the algorithms studies for minimum training error, faster convergence and achieving good performance. SLFN model parameters utilised in the proposed study are described in the Table 5.

3.4 SLFN with random hidden nodes

For N unique training data (x_i, y_i), where x_i denotes input values $[x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$ and y_i defines the target values $[y_{i1}, y_{i2}, \ldots, y_{in}]^T \in R^m$ standard SLFNs output O with $\tilde{N}$ hidden neurons, nonlinear and linear activation function $g(x)$ can be mathematically expressed as in equation (2)

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i g(w_i.x_j + b_i) = o_j \qquad j = 1, 2, \ldots, N \tag{2}$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the weight vector which connects i^{th} hidden neuron with the input neurons, $\beta_I = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{in}]^T$ defines as weight vector which connects i^{th} hidden neurons with the output neurons, and b_i defines the bias value of i^{th} hidden neuron. $w_i.x_i$ represents the dot product of w_i and x_j.

The objective of SLFN is to minimise error between actual values and output values which is defined by $\sum_{j=1}^{\tilde{N}} \|o_j - t_j\| = 0.$ There exist β_i, w_i and b_i such that

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i.x_j + b_i) = t_j, \; j = 1, 2, .., N \tag{3}$$

The above equation (3) is expressed briefly as

$$H\beta = T, \tag{4}$$

where

$$H\left(w_1, ..., w_{\tilde{N}}, b_1, ..., b_{\tilde{N}}, x_1, ..., x_N\right)$$

$$= \begin{bmatrix} f\left(a_1.x_1 + b_1\right) & \cdots & f\left(a_{\bar{N}}.x_1 + b_{\bar{N}}\right) \\ \cdots & \cdots & \cdots \\ f\left(a_1.x_N + b_1\right) & \cdots & f\left(a_{\bar{N}}.x_N + b_{\bar{N}}\right) \end{bmatrix}_{N \times \bar{N}}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\bar{N}}^T \end{bmatrix}_{\bar{N} \times m} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}$$

H can be represented as output matrix of the hidden layer. The i^{th} column of H represents the i^{th} hidden neuron output in relation to inputs. $x_1, x_2, \ldots, x_N$.

Algorithm ELM

Given a training set $\aleph = \{(x_i, t_i) \mid x_i \in R^n, t_i \in R^m \; i = 1, 2, \ldots, N\}$ activation function $g(x)$ hidden node number $\hat{N}$

Step 1 Randomly assign input weight w_i and bias b_i $i = 1, 2, \ldots, \hat{N}$.

Step 2 Calculate the hidden layer output matrix H.

Step 3 Calculate the output weight β.

$$\beta = H^{+}T$$

where $T = \{t_1, t_2, \ldots, t_N\}^T$.

3.5 Linear and nonlinear activation functions

The experiment of learning algorithm is carried out by *sklearn_extensions library* for *ELMClassifier*. According to popular belief, the ELM training algorithm can be used to train SLFNs with a number of different linear and nonlinear activation functions. The linear and nonlinear activation functions are hyperbolic tangent function (elm_tanh), relu transfer function (elm_relu), sigmoid transfer function (elm_sigmoid), sine-squared function (elm_sinsq). SLFNs has potential to implement any prediction application and estimates any continuous function. ELM provides a unified platform with a wide range of functionality of mapping techniques features. By optimising the number of hidden neurons as well as the activation function, it can be applied directly in function optimisation and multiclass applications. The soil fertility prediction problem of Davangere region are implemented by ELM with linear and nonlinear activation functions. Performance of each linear and nonlinear activation is measured and compared with other transfer functions. Table 6 describes linear and nonlinear activation functions used in the proposed study.

Table 6 Transfer functions used in the proposed study

Name of the activation function proposed	*Functionality*
reLU	$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$
Sine squared	$f(x) = \begin{cases} 1 & \text{for } x = 0 \\ \dfrac{\sin(x)}{x} & \text{for } x \neq 0 \end{cases}$
Tanh or hyperbolic tangent	$f(x) = \tanh(x)\dfrac{2}{1+e^{-2x}} - 1$
Sigmoid	$f(x) = \dfrac{1}{1+e^{-z}}$

3.6 Performance evaluation

The measure of classification performance is the accuracy score, which is calculated by multiplying the number of correct predictions with total number of predictions. The percentage of accurate predictions among $k_{samples}$ is defined as shown in equation (5).

$$accuracy(Y, y) = \frac{1}{k_{samples}} \sum_{i=0}^{k_{ssamples}-1} 1(y_i - Y_i) \tag{5}$$

where y_i is the i^{th} sample's projected value, Y_i is the real value. Kappa coefficient and AUC are the other two variables used to calculate classification and prediction performance. The kappa coefficient is described as in equation (6).

$$kappa(K) = \frac{p_o - p_e}{1 - p_e} \tag{6}$$

where p_o is the of a correct classification *possibility* and p_e is the *probability* of a correct classification. The AUC values are represented in terms of percentage of false positives. The AUC is defined in equation (7).

$$FP = \frac{FP}{(TN + FP)} \tag{7}$$

where

TP true positive

FP false positive

TN true negative

FN false negative.

4 Results and discussion

In recent years, Davangere region is highly affected by soil erosion and flood, which has declined content of nutrients in the soil. Predicting soil nutrients status in the Davangere region helps farmers in reducing the use of chemical fertilisers. Dataset of various soil nutrients are collected from soil health dashboard which is accessible publicly. To achieve zero mean and one standard deviation collected dataset is pre-processed using standard scaler and dataset is checked for the presence of NULL values. Dataset contains soil fertility attributes of primary, secondary and micro-nutrients of Davangere region. Prediction of soil nutrient deficiency is done using ELM model by changing various linear and nonlinear activation functions, i.e., tanh, reLU, sigmoid and sine squared. The performance of ELM model is assessed in terms of accuracy, kappa score, and AUC. The dataset which has been examined in the proposed study is shown in Figure 4. Samples distribution of soil nutrients such as boron, zinc, pH, and organic carbon are shown in Figure 5.

Figure 4 Soil nutrients attributes used in the proposed study

Sr.No.	Sample Nc	Latitude	Longitude	pH	EC	OC	N	P	K	S	Zn	Fe	Cu	Mn	B
1	606558/2(	14.4644	75.9218	8.1	0.7	0.82	176.6	29.35	279.6	23.56	0.49	12.35	0.96	16.58	0.52
2	606580/2(	14.4644	75.9218	8.98	0.98	0.61	275.98	28.35	406.2	24.56	0.4	13.52	0.85	15.36	0.63
3	606704/2(	14.4644	75.9218	8.59	0.68	0.2	275.65	26.53	312.52	25.36	0.63	12.14	0.43	13.25	0.52
4	KA606246/	14.4644	75.9218	7.5	0.23	0.4	278	92.43	505.28	93.8	1.76	14.48	0.93	50.58	0.3
5	KA606278/	14.4644	75.9218	8	0.39	0.7	270.09	93.71	99.52	41.4	0.25	7.69	1.19	12.24	0.5
6	KA606286/	14.4644	75.9218	6.6	0.31	1.1	180.24	69.3	255.66	50.32	1.4	59.58	3.21	33.22	0.5
7	KA606367/	14.4644	75.9218	6.8	0.23	0.8	100.24	50.06	458.2	74.6	0.3	50.36	1.43	19.09	0.5
8	KA606367/	14.4644	75.9218	6.2	0.55	1	125.28	75.3	586.4	41.6	5.64	12.02	1.43	48.2	50
9	KA606369/	14.4644	75.9218	7.5	0.26	0.5	160.34	115.4	276.47	73.9	0.12	1.62	1.7	34.64	0.5
10	KA606369/	14.4644	75.9218	7.2	0.11	0.2	180.39	185.6	255.91	63.4	0.11	5.71	0.37	13.94	0.2
11	KA606369/	14.4644	75.9218	7	0.17	0.2	120.28	121.9	466.96	62.8	0.28	1.52	0.88	27.8	0.7
12	KA606469/	14.4644	75.9218	6.2	0.27	0.3	90.25	72.59	321.12	45.4	0.12	9.47	1.43	17.75	0.4
13	KA606542/	14.4644	75.9218	7.1	0.27	0.8	120.58	97.22	350.96	60	1.01	9.25	1.1	8.08	0.1
14	KA606542/	14.4644	75.9218	7.1	0.37	1.2	140.69	101.25	352.98	92.5	0.31	4.16	0.84	14.98	0.1
15	KA606542/	14.4644	75.9218	7.3	0.25	1.2	170.89	124.4	380.35	52.5	0.27	6.94	0.9	13.04	0.1
16	KA606542/	14.4644	75.9218	6.9	0.23	1.9	140.25	60.45	320	40	0.29	8.25	1.1	13.35	0.5
17	KA606542/	14.4644	75.9218	6.8	0.33	1.9	165.89	94.7	390.58	10	0.95	2.88	0.86	8.08	0.4
18	KA606542/	14.4644	75.9218	7.3	0.38	1.4	187.25	93.17	217	52.5	0.93	1.65	0.56	5.43	0.1
19	KA606542/	14.4644	75.9218	6.5	0.51	0.9	160.36	91.17	380.19	37.5	0.99	5.66	1.2	16.26	0.2

Data about the classification task and samples distribution for each class are shown in Table 7 Davangere region soil nutrients boron and zinc cover almost equal samples for sufficient and deficient classes. Organic carbon covers high number of samples in low fertility class. pH covers almost equal distribution of samples in every class. In pre-processing step dataset is standardised to zero mean and standard deviation equal to 1.

Predicting type of pH and predicting whether soil has sufficient boron, organic carbon and zinc nutrients helps farmers in reducing excess usage of fertilisation. The cross-validation results of soil fertility classification problem are plotted in Figures 6–7. Optimal number of neurons in Figures 6–7 are identified based on the peak values in each of the graph. It can be seen in Figures 6–7 that as the number of neurons increases the training accuracy of the ELM model increases, i.e., 250 is identified as the optimal number of hidden neurons for boron nutrient and pH classification. 250 and 300 are the optimal number of neurons for zinc and organic carbon classification.

Figure 5 Evaluation of the dataset distribution

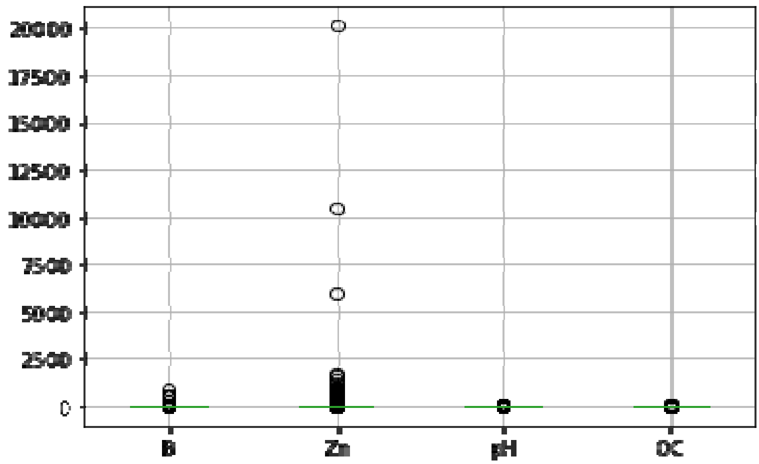


Table 7 Samples per class used in the dataset

Soil fertility	*Classes*	*Cultivation lands per class*
OC	Very low	3,200
	Low	6,900
	Moderate	3,825
	High	2,479
	Very high	1,209
B	Sufficient	8,118
	Deficient	9,495
Zn	Sufficient	9,985
	Deficient	7,625
pH	Strongly acidic	3,885
	Highly acidic	3,899
	Moderately acidic	2,485
	Slightly acidic	3,254
	Neutral	4,090

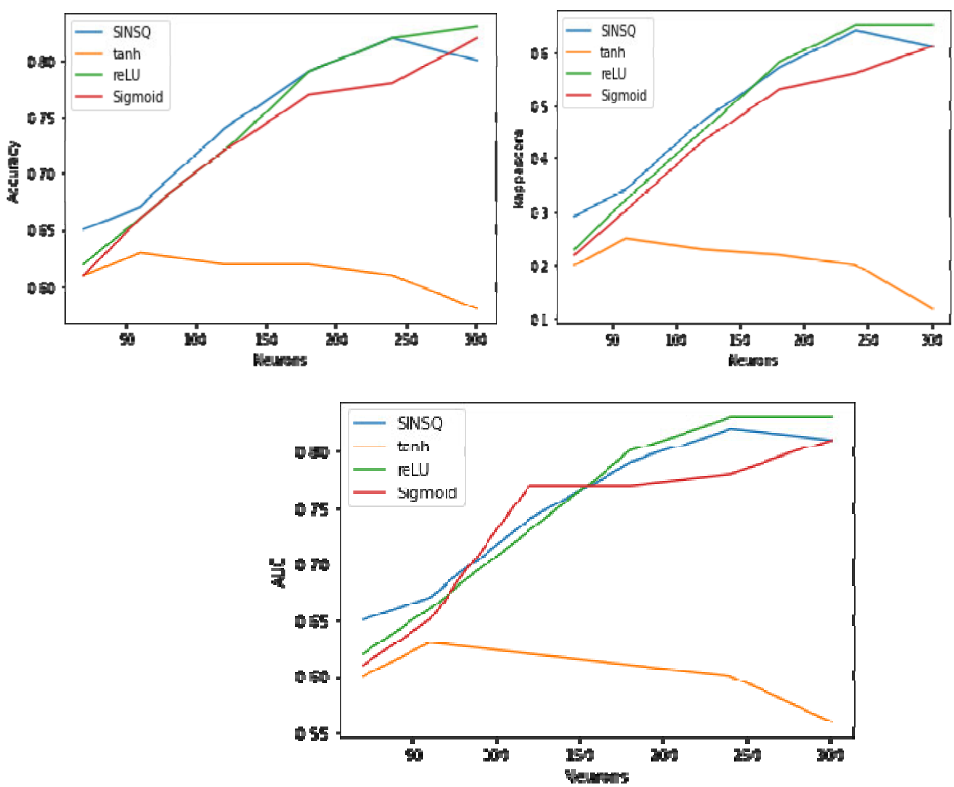
Table 8 Accuracy, kappa score and area under curve score using Sinsq activation function in boron classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
Boron	Sinsq	20	0.65	0.29	0.65
		60	0.67	0.34	0.67
		120	0.74	0.47	0.74
		180	0.79	0.57	0.79
		240	0.82	0.64	0.82
		300	0.80	0.61	0.81

Table 9 Accuracy, kappa score and area under curve score using tanh activation function in boron classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
Boron	Tanh	20	0.61	0.20	0.60
		60	0.63	0.25	0.63
		120	0.62	0.23	0.62
		180	0.62	0.22	0.61
		240	0.61	0.20	0.60
		300	0.58	0.12	0.56

Figure 6 Comparison of linear and nonlinear activation function in boron classification problem (see online version for colours)



The accuracy of the different classification problems using various transfer functions or activation functions is shown in Tables 8–23. reLU activation function has achieved best performance for boron and zinc classification problem. Sinsq nonlinear activation function has achieved best performance for organic carbon classification problem. Tanh and reLU activation functions achieved best performance for pH classification problem. Even though organic carbon classification using ELM has achieved good performance in terms of accuracy, AUC and kappa score values are very poor because classes are unbalanced. This is described in Tables 16 to 19 and in Figure 8.

Table 10 Accuracy, kappa score and area under curve score using reLU activation function in boron classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
Boron	reLU	20	0.62	0.23	0.62
		60	0.66	0.32	0.66
		120	0.72	0.450	0.73
		180	0.79	0.58	0.80
		240	0.82	0.65	0.83
		300	0.83	0.65	0.83

Table 11 Accuracy, kappa score and area under curve score using sigmoid activation function in boron classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa Score*	*Area under curve*
Boron	Sigmoid	20	0.61	0.22	0.61
		60	0.66	0.30	0.65
		120	0.72	0.43	0.77
		180	0.77	0.53	0.77
		240	0.78	0.56	0.78
		300	0.82	0.61	0.81

Table 12 Accuracy, kappa score and area under curve score using Sinsq activation function in zinc classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa Score*	*Area under curve*
Zinc	Sinsq	20	0.54	0.12	0.54
		60	0.58	0.15	0.60
		120	0.62	0.22	0.66
		180	0.65	0.28	0.70
		240	0.72	0.46	0.76
		300	0.80	0.62	0.84

Table 13 Accuracy, kappa score and area under curve score using tanh activation function in zinc classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
Zinc	Tanh	20	0.48	0.10	0.46
		60	0.49	0.13	0.48
		120	0.51	0.14	0.52
		180	0.56	0.21	0.54
		240	0.62	0.28	0.60
		300	0.68	0.34	0.65

Table 14 Accuracy, kappa score and area under curve score using reLU activation function in zinc classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
Zinc	reLU	20	0.68	0.36	0.65
		60	0.72	0.42	0.73
		120	0.78	0.50	0.79
		180	0.84	0.58	0.82
		240	0.88	0.66	0.86
		300	0.90	0.74	0.89

Table 15 Accuracy, kappa score and area under curve score using sigmoid activation function in zinc classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
Zinc	Sigmoid	20	0.50	0.26	0.52
		60	0.58	0.30	0.59
		120	0.62	0.36	0.61
		180	0.69	0.39	0.65
		240	0.66	0.42	0.67
		300	0.72	0.50	0.73

Table 16 Accuracy, kappa score and area under curve score using Sinsq activation function in organic carbon classification problem.

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
Organic-carbon	Sinsq	20	0.72	0.20	0.48
		60	0.76	0.21	0.51
		120	0.82	0.24	0.52
		180	0.85	0.25	0.51
		240	0.86	0.28	0.53
		300	0.89	0.29	0.54

Table 17 Accuracy, kappa score and area under curve score using tanh activation function in organic carbon classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
Organic-carbon	Tanh	20	0.68	0.23	0.56
		60	0.70	0.24	0.51
		120	0.71	0.21	0.55
		180	0.75	0.22	0.52
		240	0.80	0.28	0.57
		300	0.82	0.25	0.55

Table 18 Accuracy, kappa score and area under curve score using reLU activation function in organic carbon classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
Organic-carbon	reLU	20	0.65	0.10	0.48
		60	0.64	0.15	0.47
		120	0.66	0.11	0.49
		180	0.69	0.18	0.50
		240	0.72	0.20	0.46
		300	0.70	0.19	0.51

Figure 7 Comparison of linear and nonlinear activation function in zinc classification problem (see online version for colours)

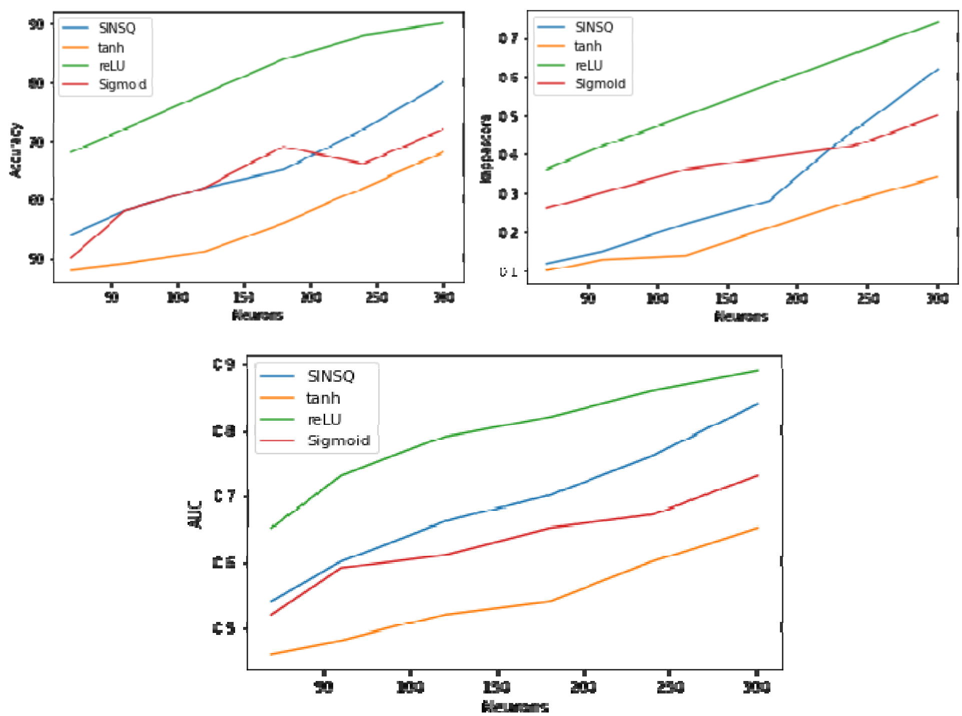


Table 19 Accuracy, kappa score and area under curve score using sigmoid activation function in organic carbon classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
Organic-carbon	Sigmoid	20	0.72	0.20	0.52
		60	0.73	0.23	0.51
		120	0.71	0.21	0.53
		180	0.78	0.28	0.55
		240	0.80	0.25	0.54
		300	0.81	0.26	0.56

Table 20 Accuracy, kappa score and area under curve score using Sinsq activation function in pH classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
pH	Sinsq	20	0.72	0.36	0.71
		60	0.76	0.42	0.78
		120	0.79	0.48	0.80
		180	0.80	0.52	0.82
		240	0.84	0.58	0.81
		300	0.88	0.64	0.89

Table 21 Accuracy, kappa score and area under curve score using tanh activation function in pH classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
pH	Tanh	20	0.78	0.48	0.80
		60	0.82	0.49	0.83
		120	0.85	0.52	0.84
		180	0.86	0.58	0.85
		240	0.89	0.64	0.90
		300	0.92	0.70	0.93

Figure 8 Comparison of linear and nonlinear activation function in organic carbon classification problem (see online version for colours)

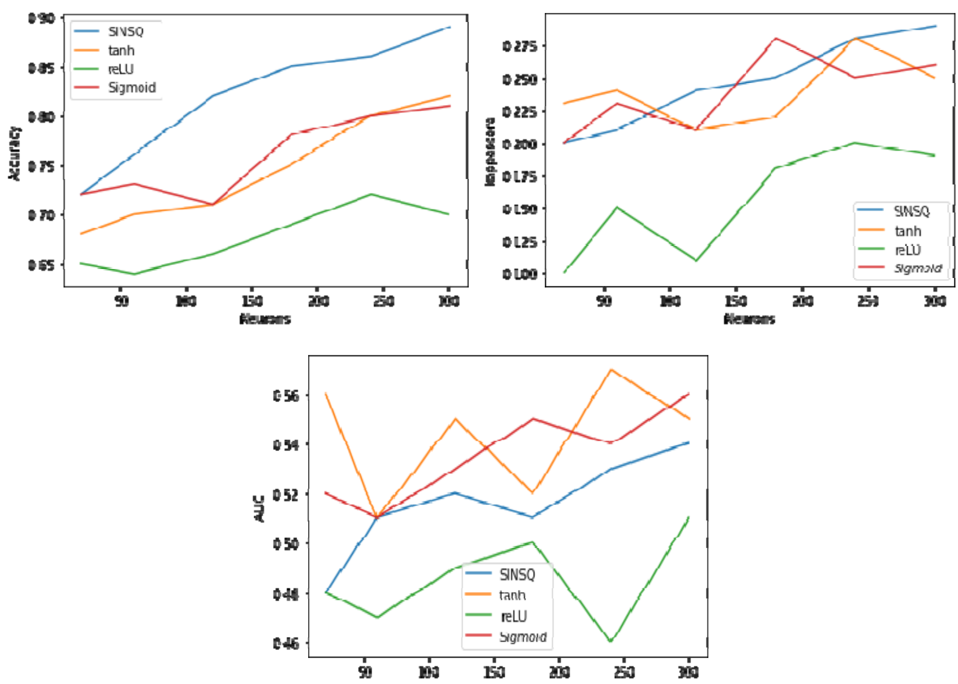


Figure 9 Comparison of linear and nonlinear activation function in pH classification problem (see online version for colours)

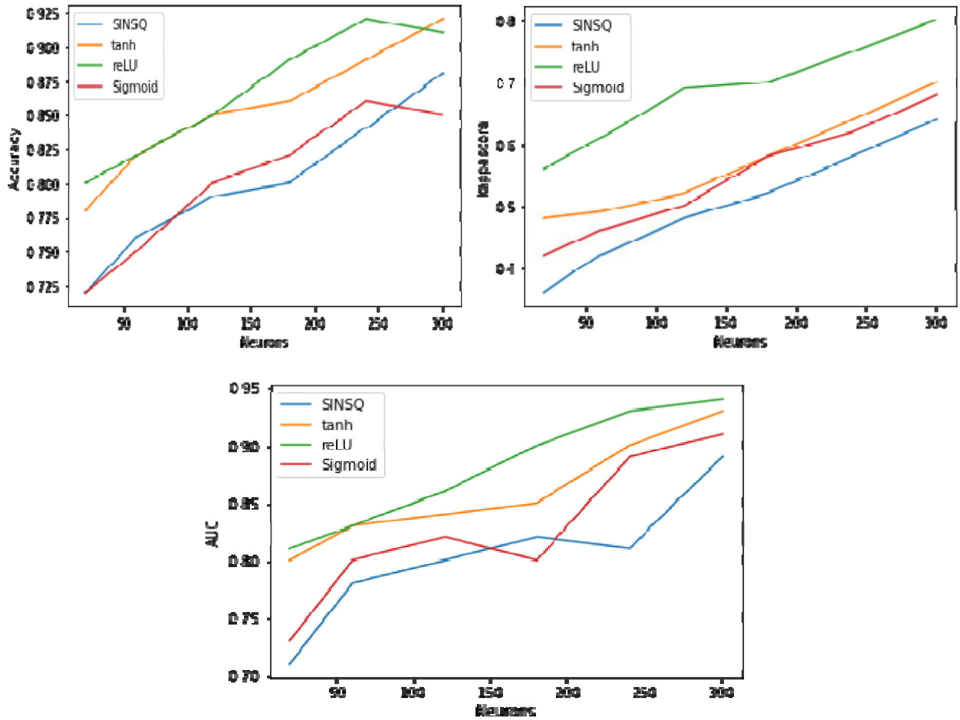


Table 22 Accuracy, kappa score and area under curve score using reLU activation function in pH classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
pH	reLU	20	0.80	0.56	0.81
		60	0.82	0.61	0.83
		120	0.85	0.69	0.86
		180	0.89	0.70	0.90
		240	0.92	0.75	0.93
		300	0.91	0.80	0.94

Table 23 Accuracy, kappa score and area under curve score using sigmoid activation function in pH classification problem

Soil fertility	*Activation function*	*Neurons*	*Accuracy*	*Kappa score*	*Area under curve*
pH	Sigmoid	20	0.72	0.42	0.73
		60	0.75	0.46	0.80
		120	0.80	0.50	0.82
		180	0.82	0.58	0.80
		240	0.86	0.62	0.89
		300	0.85	0.68	0.91

5 Conclusions

Important issue in reduction of paddy yield is due to loss in quality of soil nutrients, incorrect soil management practices and excess use of fertilisers in paddy fields. The soil test report dataset obtained from soil health dashboard created and administered by Government of India to predict soil fertility indices of organic carbon, boron, zinc and pH of Davanagere district. Fast and very efficient ELM model is used with different linear and nonlinear activation functions. The accuracy, kappa score and AUC values are used to evaluate ELM model with different activation functions. Results demonstrate that reLU activation function has achieved best performance for boron and zinc fertility problem. Sinsq nonlinear activation function has achieved best performance for organic carbon fertility problem. Tanh and reLU activation functions achieved best performance for pH problem. Results shown that optimised ELM model helps to reduce soil fertility deficiency problems of paddy crop in Davangere region.

Declaration

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by Soil Health Card Ministry of Agriculture Government of India. To digitise agriculture data, the Government of India created the soil health card. It is a massive database which has records of soil samples primary, secondary, macro and micronutrient values. Soil samples are collected from farm fields and analysed in a soil laboratory before being entered into a database.

References

de Paul Obade, V. and Lal, R. (2016) 'Towards a standard technique for soil quality assessment', *Geoderma*, March, Vol. 265, pp.96–102.

Elavarasan, D., Vincent, D.R., Sharma, V., Zomaya, A.Y. and Srinivasan, K. (2018) 'Forecasting yield by integrating agrarian factors and machine learning models: a survey', *Comput. Electron. Agric*., December, Vol. 155, pp.257–282.

Hill, M.G., Connolly, P.G., Reutemann, P. and Fletcher, D. (2014) 'The use of data mining to assist crop protection decisions on kiwifruit in New Zealand', *Comput. Electron. Agric*., October, Vol. 108, pp.250–257.

Huang, G.B., Zhou, H., Ding, X. and Zhang, R. (2012) 'Extreme learning machine for regression and multiclass classification', *IEEE Trans Syst, Man, Cybernet, Part B (Cybernetics)*, Vol. 42, No. 2, pp.513–529.

Huang, G.B., Zhu, Q.Y. and Siew, C.K. (2006) 'Extreme learning machine: theory and applications', *Neurocomputing*, Vol. 70, Nos. 1–3, pp.489–501.

Islam, T., Rizan, R.U.B. and Tusher, Y.A. (2020) 'Nitrogen fertilizer recommendation for paddies through automating the leaf color chart (LCC)', *International Journal of Advanced Computer Science and Applications*, Vol. 11, No. 8, DOI: 10.14569/IJACSA.2020.0110891.

Jia, H.Y., Chen, J., Yu, H.L. and Liu, D.Y. (2010) 'Soil fertility grading with Bayesian network transfer learning', in *International Conference on Machine Learning and Cybernetics*, IEEE, pp.1159–1163.

Jose, A. et al (2021) 'Detection and classification of nutrient deficiencies in plants using machine learning', *J. Phys.: Conf. Ser.*, No. 1850 012050.

Kodaty, S.C. and Halavath, B. (2021) 'A new approach for paddy leaf blast disease prediction using logistic regression', *Advances in Information Communication Technology and Computing, Lecture Notes in Networks and Systems*, Vol. 135, https://doi.org/10.1007/978-981-15-5421-6_51.

Krishnan, P., Venugopal, K.R. and Sehgal, J. (1996) *Soils of Kerala for Optimising Land Use*, p.48, NBSS Publication, Delhi.

Mucherino, A., Papajorgji, P. and Pardalos, P.M. (2009) 'A survey of data mining techniques applied to agriculture', *Oper. Res. Int. J.*, Vol. 9, No. 2, pp.121–140.

Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R.L. and Mouazen, A.M. (2016) 'Wheat yield prediction using machine learning and advanced sensing techniques', *Comput. Electron. Agric.*, February, Vol. 121, pp.57–65.

Reashma, S.J. and Pillai, A.S. (2017) 'Edaphic factors and crop growth using machine learning – a review', in *International Conference on Intelligent Sustainable Systems (ICISS)*, IEEE, pp.270–274.

Ritz, C., Putku, E. and Astover, A. (2015) '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.*, Vol. 66, No. 3, pp.548–554.

Romero, J.R., Roncallo, P.F., Akkiraju, P.C., Ponzoni, I., Echenique, V.C. and Carballido, J.A. (2013) 'Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires', *Comput. Electron. Agric.*, August, Vol. 96, pp.173–179.

Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappie, M., Märker M. et al (2017) '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.*, December, Vol. 601, pp.821–832.

Sethy, P.K., Barpanda, N.K., Rath, A. and Behera, S.K. (2020) 'Nitrogen deficiency prediction of rice crop based on convolutional neural network', *Journal of Ambient Intelligence and Humanized Computing*, November, DOI: 10.1007/s12652-020-01938-8.

Singh, B.K., Singh, R.P., Bisen, T. and Kharayat, S. (2018) 'Disease manifestation prediction from weather data using extreme learning machine', *IEEE Transactions*, ISBN: 978-1-5090-6785-5/18.

Sirsat, M.S., Cernadas, E., Fernández-Delgado, M. and Barro, S. (2018) 'Automatic prediction of village-wise soil fertility for several nutrients in India using a wide range of regression methods', *Comput. Electron. Agric.*, November, Vol. 154, pp.120–133.

Sirsat, M.S., Cernadas, E., Fernández-Delgado, M. and Khan, R. (2017) 'Classification of agricultural soil parameters in India', *Comput. Electron. Agric.*, Vol. 135, pp.269–279.

Terhoeven-Urselmans, T., Vagen, T.G., Spaargaren, O. and Shepherd, K.D. (2010) 'Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library', *Soil Sci. Soc. Am. J.*, Vol. 74, No. 5, pp.1792–1799.

Veronesi, F., Korfiati, A., Buffat, R. and Raubal, M. (2017) 'Assessing accuracy and geographical transferability of machine learning algorithms for wind speed modelling', *Information Science*, pp.297–310, Springer, Cham.

Wang, B., Waters, C., Orgill, S., Cowie, A., Clark, A., Liu, D.L. et al. (2018) 'Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia', *Ecol. Ind.*, May, Vol. 88, pp.425–438.