BIPARTITE QUADRATIC FISHER SCORE AND TRIPLET LOSS BASED DEEP LEARNING FOR SOIL QUALITY PREDICTION

Balaji.G[1], Dr.P.Vijayakumar[2]

[1]Ph.D. Research Scholar, Department of Computer Science, Sri Jayendra Saraswathy Maha Vidyalaya College of Arts and Science. Coimbatore, indis

[2]Associate Professor, Department of Computer Applications, KG College of Arts and Science. Coimbatore,India

Correspondence: Balaji.G; E.mail:balajiphd08@gmail.com

Abstract

The vital center of attention to the management of soil to improve crop productiveness was a conservation as well as refinement of vigorous soil. Some of reasons like, population pressures, terrestrial disadvantages as well as conventional soil mechanisms are charge of to worsening or declining in soil fertility as far as developing countries like India. Hence, a considerable improvement within crop is arrived at applying appropriate crop fitness technique. On other hand, an increase in productivity is said to be arrived at via efficient soil resource management and remedial assessments. Timely identification issues associated using soil management for ensuring productivity. Over the past few years, classification issues were efficiently employed as ML and Deep Learning (DL) techniques. Therefore, identifying the accurate and precise method to predict the soil quality is still the research of interest. Bipartite Quadratic Fisher Score as well as Triplet Loss-based Deep Learning (BQFS-TLDL) is proposed for predicting the soil quality in an accurate and precise manner. First, with the soil moisture prediction big dataset provided as input, significant and precise features are selected by means of Bipartite Quadratic Mutual Information and Fisher Score model. Next, with the significant features selected, Triplet Loss-based Deep Learning is applied for predicting soil quality. Simulation of deep learning methods, BQFS-TLDL has better accuracy as well as efficiency with minimum error rate.

Keywords: Machine Learning, Bipartite, Quadratic, Mutual Information, Fisher Score, Triplet Loss, Deep Learning, Soil Quality Prediction

DOI Number: 10.14704/NQ.2022.20.12.NQ77236 **NeuroQuantology2022;20(12): 2549-2571**

1. Introduction

With the increase in the world's population and also scarce availability of waterresources owing to several reasons like, change in climatic conditions, tough competition between numerous uses as well as inflexible agriculture, necessity to evolution of several sustainable factories.

Data driven scheme was proposed in [1] where detailed analysis was made by employing EF. EF acquired from Landsat images was utilized post validation by means of eddy covariance measurements. Also, Time Domain Reflectometry (TDR) sensors were employed for obtaining data between 2018 and 2020. Moreover, some of the predictorvariables included was of meteorological, characteristics of soil, EF. Strong correlation among soil moisture as well as RF predicted soil moisture at every sensor locations for ensuring accuracy. Despite accuracy measure, the time consumed in ensuring the soil characteristics was not focused.

Partial least square regression was proposed in [2] with the purpose of providing prediction results with respect to soil fertility and crop yield based on the historical factors. With this objective as design, data were acquired via several elements. With these factors the input period as well as analysis of corresponding soil was achieved. Finally, error analysis was made for prediction of soil fertility. Though error analysis for soil fertility prediction was made, the accuracy and time factor was not concentrated.

In order to handle such limitations, Bipartite Quadratic Fisher Score and Triplet Loss-based Deep Learning (BQFS-TLDL) method for improving precision, accuracy as well as significant feature selection. From significant and precise characteristic from the overall feature set of every predictor on Bipartite Quadratic function, is predictable to not only select significant feature but also distinguish between two types of crops, maize and peanuts for the prediction. This was performed by employing two functions, mutual information and fisher score. And then, a deep learning based long short term memory modulated by updated weight based on Triplet Loss function was designed. The deep learning is developed for describe vectors across every predictors by predicting the soil quality in an accurate manner. We tested our BQFS-TLDLmethod using soil moisture prediction big dataset.

1.1 Contributory remarks

The contributions of the proposed Bipartite Quadratic Fisher Score and Triplet Loss-based Deep Learning (BQFS-TLDL) are listed below:

- A novel Triplet Loss-based LSTM for enhancing soil quality prediction.
- Bipartite Quadratic form employing Mutual Information Coefficient (MIC) and Fisher Score (FS) to summarize significant and precise feature from the overall feature set and model the soil quality prediction.
- To propose Triplet Loss-based Deep Learning by learning complicated association between features by employingLSTM and integrating Triple Loss while updating weights, therefore ensuring accurate prediction.
- Extensive experimental evaluation of BQFS-TLDL method against Random forest ensemble and Partial least square regression methods.

1.2 Organization of the paper

The article is summarized by. Section 2 reviews related work. Section 3 describes proposed Bipartite Quadratic Fisher Score and Triplet Loss-based Deep Learning (BQFS-TLDL) method. Section 4 and 5, explains the experiment results and discussion. Finally, Section 6 provides concluding remarks.

2. Related works

Soil classifications as well as environment necessitate field or laboratory testing of certain amount of fundamental characteristics as well as several constituent factors. For this specific purpose, simple field mechanisms are accessible for this intent with the aid of hydrochloric acid dissolution.

However, as far as the land assessment and classification is concerned, outburst or outbreak to a laboratory is short that in turn causes restricted analysis of laboratory and therefore result in classification errors. Hence, bestowing a swift and reasonable substitute would result in soil resource management.

An elaborate comparative analysis of distinct machine learning methods employing LUCAS was proposed in [3]. In this work, three types of mechanisms for predicting features of varied mineral soil were tested. An in-depth analysis concerning clay content, SOC and N were also analyzed.

Over the past few years, scientists have been developing more effective mechanisms to enhance spatial distribution soil prediction, however, the in this field is said to be limited. An algorithm was introduced in [4] to obtain terrain factors at multiple scales and also calibrated in an automatic manner for each predictor by employing RF, for enhancing accuracy.

Owing to the drastic change observed in climatic conditions, deficiency is identified by acute menaces as due to the lack of water being identified as one of the most restricting characteristics for plant growth. LSTM [5] was outlined as well as pointed by employing different data periods. However, accurate prediction was not ensured. To address on this aspect, an attention-aware based LSTM was proposed in [6] to design efficient deep learning models for accurate land surface prediction.

Over the recent few years, the insistence for food on the basis of exhaustive agriculture has reduced the quality of soil, resulting in considerable issues like, growing agricultural productivity and stimulating environmental feasibility. Hence, researchers are designing mechanisms for soil quality estimation on the basis of artificial intelligence models for multidimensional data processing using agro-industrial systems. This in turn would provide essential information from the farmer side to ensure smooth mechanisms for soil cultivation in equal proportionate and in a timely manner.

Filter based instance selection was employed in [7] for soil data classification. A review of both machine learning and remote sensing techniques were investigated in [8] with the purpose of estimating the different types of soil indicators. Yet another assessment on sophisticated mechanisms employing different ML techniques was investigated in [9].

The prediction of crop yield were performed for ensuring better crop yield by employing machine learning algorithms that are said to be one of the demanding considerations as far as agricultural sector is concerned. Owing to this growing importance of crop yield prediction, a comprehensive review on the utilization of ML for predicting crop yield with exceptional significance on predicting palm oil was investigated in [10].

One of the new conceptions that make agriculture more effective and significant in smart farming is the employing of sophisticated technologies. Evolutions in automation as well as artificial intelligence empower farmers for observing every practice as well as accurate action acquired using superhuman accuracy. Upon comparison with other algorithms, deep learning technique posses the advantage of extraction of feature in an automatic manner without the intervention of human. Advantages of deep learning in agricultural aspects were concentrated in [11]. Yet another comprehensive review on machine learning for precision agriculture was proposed in [12].

To make certain higher construction of crop employing constrained land resources, it was paramount for detecting pertinent soil owing to the reason that distinct crops require different soil types. Over the past few years, two types of methods are found in determining the type of the soil, i.e., chemical and image analysis. Despite accuracy observed in determining the soil types based on chemical factors. Then, classification of soil on the images was though found to be cheaper as well as quicker accuracy was said to be less.

In [13], feature based algorithm integrates Q-HOG was proposed for classifying soil types in a precise manner. On the other hand, extreme learning machine parameters were employed in [14] for soil nutrient classification therefore ensuring accuracy to a greater extent. Though these methods were found to address the issues concerning soil quality prediction however certain inadequacies were found to persist, like, non-linear mapping among raw data with extracted feature quality.

Deep recurrent Q-network was implemented in [15] for predicting crop yield. Here, the linear layer mapped the output values to Q-values. Followed by which reinforcement learning provided fusion of parametric features using threshold within crop yield prediction by reducing error and increasing the forecast accuracy to a greater extent.

Fuzzy logic tool was employed in [16] to forecast the fertility of soil in Nigeria. In [17], soil quality identification and predication were made on the basis of traditional soil andvegetation classifications. Different modeling scenarios were proposed in [18] using artificial intelligence models for analyzing the role of ground water quality in soil prediction. However, with weather to be unpredictable with the growing season agricultural breeding was compromised. To address on this issue, Support Vector Regression with Radial Basis Kernel Function was introduced in [19] for better yield prediction. In [20], an elaborate review on deep learning techniques for soil quality prediction was investigated.

Motivated by the above challenges, in this work, a novel soil quality prediction method, called, Bipartite Quadratic Fisher Score and Triplet Loss-based Deep Learning (BQFS-TLDL) is developed. The detailed explanation of BQFS-TLDL was provided below.

3. Methodology

We propose a novel Bipartite Quadratic Fisher Score and Triplet Loss-based Deep Learning (BQFS-TLDL) for soil quality. BQFS-TLDL is designed for soil quality prediction. Initially, Bipartite Quadratic Mutual Information and Fisher Score-based feature selection is presented. The second is the Triplet Loss-based Deep Learning for Soil Quality Prediction. Figure 1 shows the block diagram of BQFS-TLDL method.

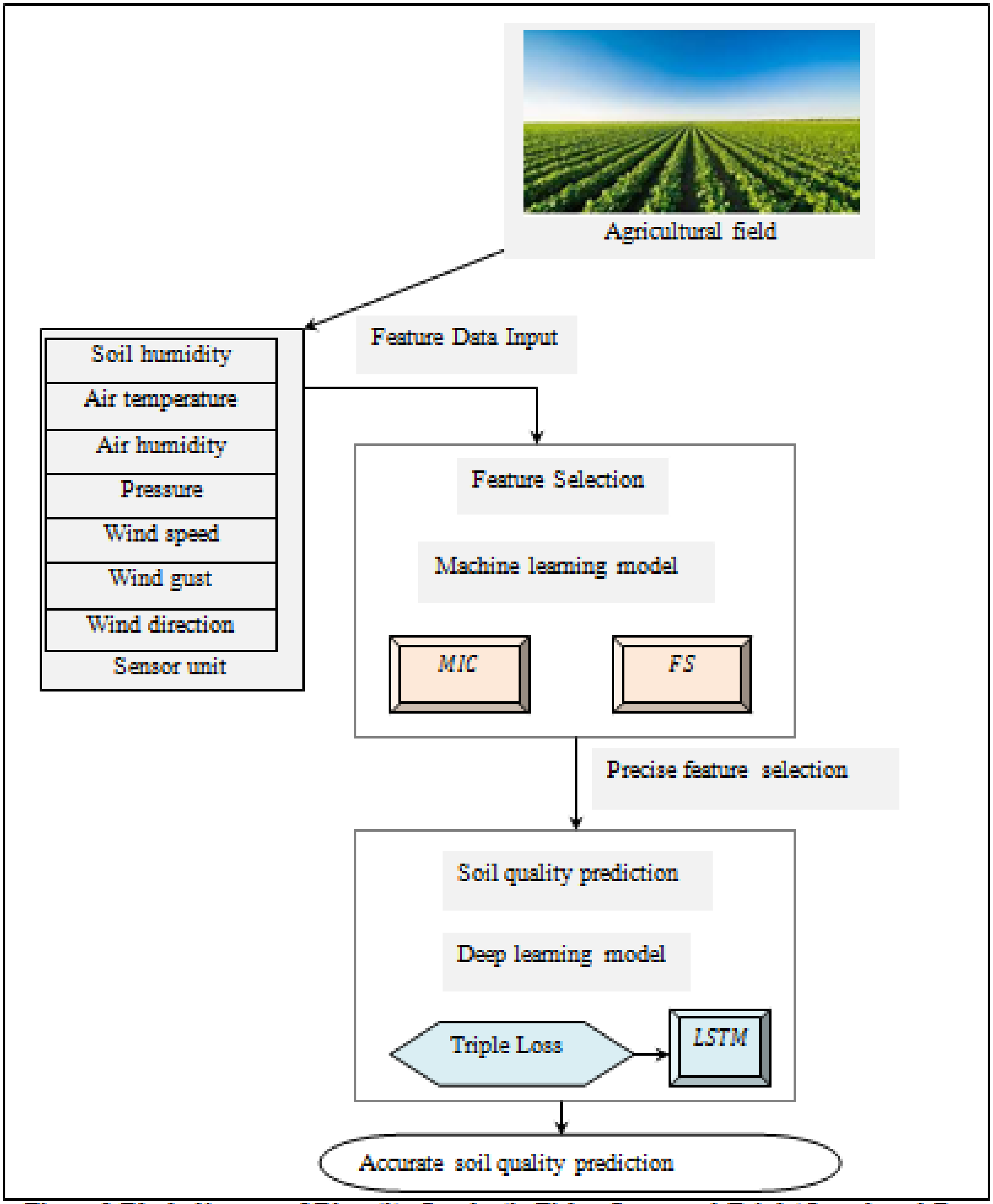

Figure 1 Block diagram of Bipartite Quadratic Fisher Score and Triplet Loss-based Deep Learning

As given in the above block diagram with the feature data input acquired from the corresponding sensor unit via soil moisture prediction big data dataset, the overall work is split into two sections, namely, feature selection and soil quality prediction. The elaborate briefing of work was narrated after the dataset description.

3.1 Dataset description

The Wazihub soil moisture prediction dataset was acquired as a segment of an experiment performed by Wazihubemploying internet of things (IoT) sensors over a period of 4 months from 4 distinct plots growing maize and peanuts in Senegal. In addition, the plots were positioned right next to each other, separated by a one meter perimeter.Also three distinct irrigation schedule types were utilized. They were, usual irrigation, less than crop needs and based on water loss respectively.

The objective of this work remains in accurate soil quality prediction multiple days in advance, owing to this assists the farmers in preparing their irrigation schedules in a more effective manner. Moreover, the fields were irrigated by cultivating crops as follows. In addition, the IoT soil moisture sensors were positioned in each of the fields following which IoT weather station was placed near the fields, where the IoT devices transmitted following data in five minute intervals. The transmitted data are:

Soil humidity
Air temperature (C)
Air humidity (%)
Pressure (KPa)
Wind speed (Km/h)
Wind gust (Km/h)
Wind direction (Deg)

Some of other data acquired by hand on a daily basis are as given below:

- Minimumdaily temperature
- Maximum daily temperature
- Relative air humidity
- Wind speed
- Solar Irradiance
- Radiant energy emitted by the Sun
- Crop coefficient
- Evapotranspiration rate observed in the crop

- Evapotranspiration rate observed for a well calibrated reference crop
- Rainfall per day
- Water need

The entire dataset is split into four different tables, namely, SampleSubmission, Train, ContextDataMaize and ContextDataPeanut (i.e., context data). They are listed as below. The training data contains the soil humidities for 4 distinct fields acquired every five minutes via IoT.

Table 1 Training data file

S. No	Feature	Description
1	Timestamp	Time of recording
2	Soil_Humidity_1	Soil humidity of field 1
3	Irrigation_Field_1	1 = irrigation on; 0 = irrigation off
4	Soil_Humidity_2	Soil humidity of field 2
5	Irrigation_Field_2	1 = irrigation on; 0 = irrigation off
6	Soil_Humidity_3	Soil humidity of field 3
7	Irrigation_Field_3	1 = irrigation on; 0 = irrigation off
8	Soil_Humidity_4	Soil humidity of field 4
9	Irrigation_Field_4	1 = irrigation on; 0 = irrigation off
10	Air_Temperature	Temperature of the air
11	Air_Humidity	Humidity of the air
12	Air_Pressure	Pressure of the air
13	Wind_speed	Speed of the wind
14	Wind_Gust	Gust of the wind
15	Wind_Direction	Direction of the wind

Second, the context data collected for both maize and peanuts are given below.

Table 2 Context data collected

S. No	Feature	Description
1	Date	Date
2	Min_Temp	Minimum temperature
3	Max_Temp	Maximum temperature
4	Humidity	Present air humidity
5	Wind_Speed	Wind speed
6	Solar_Irradiance	The power unit area
7	Sun	Radiant energy emitted by the sun
8	Kc	Crop coefficient
9	ETc	Evaporation rate observed in crop
10	ETo	Evaporation rate observed in reference crop
11	Rainfall	Rainfall per day
12	Water_need_1_day	Water needs of the crop per day
13	Water_need_2_days	Water needs of the crop two times a day
14	Water_need_3_days	Water needs of the crop three times a day

Finally, the sample submission, an example providing what the actual submission will look like is given below.

Table 3 Sample submission

S. No	Feature	Description
1	ID	Date and time
2	X	X
3	Name of each field	Soil humidity 1, soil humidity 2, soil humidity 3 or soil humidity 4
4	Values	Prediction result

3.2 Bipartite Quadratic Mutual Information and Fisher Score-based feature selection

The Wazihub soil moisture prediction dataset employed in our work contain redundant information that harms the further soil quality prediction task. Feature selection (FS) is referred to as pre-processing technique, wherein feature subsets are selected from the given input dataset that contain significant information than total feature set. Feature selection is therefore considered as the challenging issue and consists of huge attributes. From overwhelming popularization of ML, searching process to identify the subset of optimal features is yet under study.

In this work, combinatorial approach as a Bipartite Quadratic form employing Mutual Information Coefficient (MIC) and Fisher Score (FS) in application to the underlying relevant feature selection problem for soil quality prediction.With the proposed Bipartite Quadratic-based Mutual Information Coefficient (MIC) and Fisher Score (FS) significant feature selection, training is said to be faster, reduces the overall complexity and hence maximizes the accuracy when selecting the right subset. Figure 2 shows the block diagram of Bipartite Quadratic Mutual Information and Fisher Score-based feature selection.

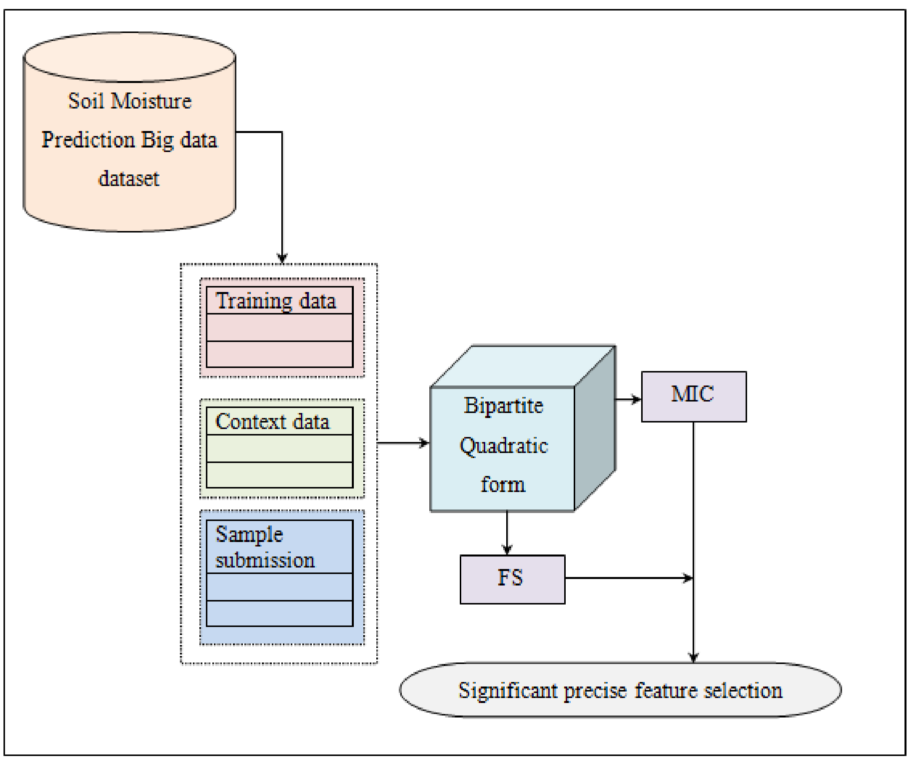

Figure 2 Bipartite Quadratic Mutual Information and Fisher Score-based feature selection

From figure 2, given a Dataset '$DS\ \{(p_i, q_i)\}_{i=1}^{n}$' with Features '$F = f_1, f_2, \dots, f_n$' (i.e., soil quality characteristics like, time, temperature, humidity, wind speed and so on), where '$p_i \in R^m$' and '$q_i \in \{1,2, \dots, k\}$', the objective in our work remains in identifying a feature subset of size 'k' that consists of the most significant features. Let, '$P = [p_1, p_2, \dots, p_n] \in R^{m*n}$' represent the data matrix, with 'p^i' denoting the '$i - th\ row$' of 'P', '1' denoting the vector of all ones, '0' denoting the vector of all zeros for the corresponding identity matrix 'I' respectively.

In order to compare between the features (i.e., soil quality characteristics) from a given feature subset, the distance function has to be defined and hence the binary optimization problem is formulated. For that initially, the mutual information between the features in the feature set has to be obtained and expressed by,

$$MI(p, q) = \sum_{p,q} Prob(p, q) \ln \frac{Prob(p,q)}{Prob(p)Prob(q)} \quad (1)$$

From the above equation (1), 'p', 'q' represents the sample features in consideration for obtaining or retrieving mutually informative feature. In our case we are looking at mixes of two types of crops maize as well as peanuts to estimate soil quality prediction. Numerous models are available for binning two distinct types (i.e., crops). In our work probability measure is utilized for binning mutually informative feature, with each bin being each interval, the '$Binary$' binning is mathematically represented as given below.

$$I(p, q) = \psi(m) - < \psi(p) > + \ \psi\left(Dis(p, q)\right) - < \psi\ (nneigh) > \quad (2)$$

From the above equation (2), 'ψ' denotes the digamma function, '$Dis(p, q)$' the distance between the features and the designated nearest neighboring feature and '$nneigh$' denotes the number of neighbors. Then, with the above information criterion, the binary binning matrix was expressed by,

$$C(p, q)_{i,j} = \frac{I(p,q)_{i,j}}{\log 2Min\ (p,q)} \quad (3)$$

From the above equation (3), '$I(p, q)_{i,j}$', represent binned values of 'p, q'. Finally, with this formulation, the quadratic matrix 'Q_{ij}' that ensures minimum redundancy is obtained. We want to identify feature combinations that are not correlated. In other words, the identified features should be more distant from each other, but maintaining relevancy for soil quality prediction. This quadratic matrix is mathematically formulated for two different types of crops as given below.

$$Q_{ij} = \begin{cases} if\ (i = j);\ -|Dis(P_i, q); C = Maize| \\ if\ (i < j); \left|Dis\left(P_i, P_j\right); C = Peanuts\right| \end{cases} \quad (4)$$

From the above equation (4), the feature combinations (i.e., for both maize and peanuts) are established. By taking the absolute value of distance function, higher weight to features possessing higher relevancy are provided and vice versa for feature combinations. Then, the fisher score for the above quadratic matrix 'Q_{ij}' to identify subset of relevant feature is mathematically expressed as given below.

$$F(Q_{ij}) = \{(S_B)(S_T + \alpha I)^{-1}\} \quad (5)$$

From the above equation (5), 'S_B' and 'S_T' denotes the between and total scatter matrix, with 'α' representing the regularization parameter for the identity matrix 'I' respectively.

$$S_B = \sum_{k=1}^{C} Size_k (\mu_k - \mu)(\mu_k - \mu)^T \quad (6)$$

$$S_T = \sum_{i,j=1}^{m,n} (Q_{ij} - \mu)(Q_{ij} - \mu)^T \quad (7)$$

From the above equations (6) and (7), 'μ_k' and '$Size_k$' forms the mean vector and the size of the '$k - th\, class$' (i.e., either maize of peanuts) in the reduced quadratic form 'Q_{ij}'. Finally, the relevant features selected are as given below.

$$(FS) = MAX\,\{F(Q_{ij})\} \quad (8)$$

As given from the above equation (8), by finding the maximal feature values using quadratic matrix, relevant features are selected. The algorithmic process of Bipartite Quadratic-based feature selection was expressed as.

Input: Dataset 'DS', Features '$F = f_1, f_2, \ldots, f_n$'
Output: Precise and relevant features
1: **Initialize** '$n = 14$', sample features 'p', 'q', number of neighbors '$nneigh$' 2: **Initialize** identity matrix 'I' 3: **Begin** 4: **For** each Dataset 'DS' with Features 'F' 5: Evaluate mutual information between features as in equation (1) 6: Formulate binary binning as in equation (2) 7: Estimate binary binning matrix as in equation (3) 8: Obtain quadratic matrix as in equation (4) to retrieve feature combinations (i.e., for maize and peanuts) 9: **For** each quadratic matrix 'Q_{ij}' 10: Evaluate the fisher score as in equation (5) 11: Obtain features selected as in equation (8) 12: **Return** feature selected 'FS' 13: **End for** 14: **End for** 15: **End**

Algorithm 1 Bipartite Quadratic-based feature selection

From Algorithm 1 Bipartite Quadratic-based feature selection algorithm, needs in returning precise and significant features that in turn forms the basis for soil quality prediction. With this objective a bipartite quadratic form is designed where the bipartite here represents the formulation of two types of crops, i.e., maize, peanuts and quadratic denotes the significant features to be selected separately and position it in two different quadrants. First, mutual information between features are obtained via binning matrix. Second to retrieve feature combinations quadratic matrix is designed. Finally, fisher score is utilized to select the precise features among the feature set.

3.3 Triplet Loss-based Deep Learning for Soil Quality Prediction

Initially soil quality necessitates not only current or prevailing data. Compared with traditional quantum assisted prediction algorithms, LSTM in all aspects examines time. With aid of LSTM, soil quality analysis ranging between time series forecasting predict future prediction results on the basis of previous, sequential data. This in turn ensures soil quality prediction accuracy that achieves better decision making for the farmers. Despite the learning of complicated association between features by employing LSTM, owing to the learning process highly prone according to time, the error or false positive rate is also said to be compromised. To address on this aspect, Triplet Loss-based Deep Learning for Soil Quality Prediction and applying it to LSTM was developed. Figure 3 illustrates Triplet Loss-based Deep Learning for Soil Quality Prediction.

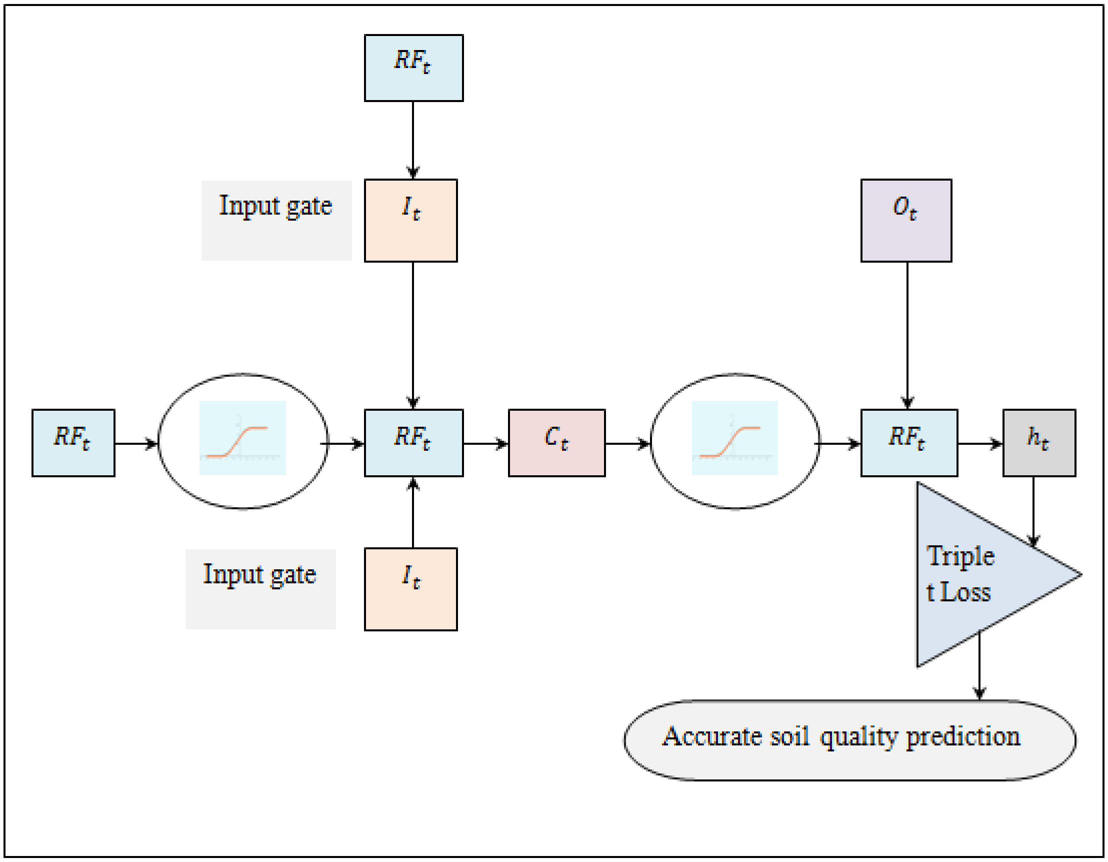

Figure 3 Triplet Loss-based Deep Learning for Soil Quality Prediction

As shown in the above figure, the Triplet Loss-based Deep Learning for Soil Quality Predictionusing LSTM consists of four layer structure with interactions between four layer structures. Here 'h_{t-1}' and 'h_t' denote the output of the previous cell and the current cell, '$I_t = RF_t$' denotes the input, soil state represented by 'C_t' at time 't', and 'F_t' denotes the forget gate that controls the probability of forgetting via the sigmoid activation function 'σ' and 'O_t' denotes the output state to get the finaloutput by employing triplet loss function.

Triplet loss is the reference input (termed as anchor or the soil sample provided as input with the significant features selected) is compared to a matching input (called positive samples or correctly predicted soil) as well as non-matching input (termed as negative samples or incorrectly prediction soil). Here, the interval among anchor as well as positive samples was reduced, and the interval between the anchor and negative samples is maximized. With application of Triplet loss for optimizing the soil quality predictionresults, that is applied for fine tuning the weight parameters, including anchor, positive and negative samples, that make the interval between anchor and positive samples to be smaller than the interval between anchor and negative samples.

Initially, with the significant features selected for the corresponding soil dataset, the candidate storage is mathematically stated as given below.

$$c_t = \tanh(W_c[h_{t-1}, x_t] + b_c), x_t \in RF \quad (9)$$
$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (10)$$
$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (11)$$
$$C_t = f_t * C_{t-1} + i_t * c_t \quad (12)$$
$$O_t = \sigma(W_O[h_{t-1}, x_t] + b_O) \quad (13)$$
$$h_t = O_t * \tanh(C_t) \quad (14)$$

From the above equations (9), (10), (11), (12), (13) and (14), 'W_c', 'W_i', 'W_f' and 'W_O' denotes the weight coefficient corresponding to candidate storage unit, input unit, forget unit and output unit, 'b_c', 'b_i', 'b_f' and 'b_O' denoting the biasfor each predictor 'h_{t-1}' with the resultant values stored in the candidate storage state 'c_t', input gate 'i_t', forget gate 'f_t', current cell state 'C_t' and output gate 'O_t' at time stamp 't' respectively.

Finally, the two activation functions employed are 'σ' and 'tanh' respectively. The value of the input gate 'i_t' is obtained by updating current input data (i.e., current soil state sample features) candidate state (i.e., overall soil state sample features). On the other hand, the forget gate 'f_t' controls update (i.e., the previous soil state sample features) to storage part (i.e., sample soil state features acquired from the storage unit).

Next, the current cell state 'C_t' is obtained on the basis of the sate value of last unit 'C_{t-1}'. Followed by which, the output gate 'O_t' controls storage unit with which output unit results are obtained in 'h_t'. Finally, with the aid of the Triplet Loss function that is being employed as a significant learning indicator for optimizing the soil quality predictive that is applied for fine tuning the weight parameters of predictivemodel.

$$TLF = \sum_{i=1}^{n} \left[\left| f(h_i[t]) - f\left(h_i^p[t]\right) \right|_2^2 - \left| f(h_i[t]) - f(h_i^n[t]) \right|_2^2 \right] \quad (15)$$

From the above equation (), the Triplet Loss function 'TLF' is estimated for the hidden unit at time 't' by means of the hidden unit at time 't' with respect to positive samples '$h_i^p[t]$' and hidden unit at time 't' with respect to negative samples '$h_i^n[t]$' respectively. By means of estimating this loss function, the falsification of soil quality prediction can be reduced to a greater extent. The pseudo code representation of Triplet Loss-based Deep Learning for Soil Quality Prediction is given below.

Input: Dataset 'DS', Features '$F = f_1, f_2, \ldots, f_n$'
Output: Accurate soil quality prediction
1: **Initialize** relevant features 'RF' 2: **Initialize** weight coefficient corresponding to candidate storage unit 'W_c', input unit 'W_i', forget unit 'W_f', output unit 'W_o' 3: **Initialize** bias for candidate storage unit 'b_c', input bias 'b_i', forget bias 'b_f', output bias 'b_o' 4: **Begin** 5: **For** each Dataset 'DS' with Features 'F' 6: Estimate candidate storage unit as in equation (9) 7: Estimate the value of input gate as in equation (10) 8: Obtain the value of forget gate as in equation (11) 9: Obtain current cell state as in equation (12) 10: Evaluate output gate as in equation (13) 11: Evaluate output unit as in equation (14) 12: Estimate Triplet Loss function as in equation (15) 13: **End for** 14: **End**

Algorithm 2 Triplet Loss-based Deep Learning for Soil Quality Prediction

As given in the above Triplet Loss-based Deep Learning with the objective of improving the accuracy of soil quality prediction in a faster manner with minimum falsification, a Triplet Loss function was introduced while updating weights between layers in LSTM. By employing this Triplet Loss function for optimizing soil quality predictiveapplied for fine tuning the weight parameters of predictivemodel, accurate prediction is ensured.

4. Experimental setup

The proposed Bipartite Quadratic Fisher Score and Triplet Loss-based Deep Learning (BQFS-TLDL) method was implemented by various methods, such as Random forest ensemble[1] and Partial least square regression[2]. A persuading characteristic to assessment parameter was potentiality for differentiating among dl methods developed in Python using the by Soil Moisture Prediction dataset (https://zindi.africa/competitions/wazihub-soil-moisture-prediction-challenge/data).The effectiveness of the learning method was obtained for estimating method numerous execution. BQFS-TLDL method is examined as,

- Soil quality prediction time
- Soil quality prediction accuracy
- Precision
- False positive rate

5. Discussion

5.1 Case scenario 1: Soil quality prediction time

In this section, soil quality prediction time is estimated. Faster or earlier the soil quality prediction is made, accuracy crop yield to the user of farmer is said to be ensured. Hence, soil quality prediction time plays a major role in analyzing the soil quality. It is calculated as given below.

$$P_{time} = \sum_{i=1}^{n} S_i * Time\ [SQP] \quad (16)$$

In (16), soil quality prediction time 'P_{time}' is calculated on 'S_i' collected from four different plots of land that were planted with either maize or peanuts as well as actual time consumed within quality '$Time\ [SQP]$' via deep learning. Table 4 given below lists the soil quality prediction time for three different methods, BQFS-TLDL, Random forest ensemble [1] and Partial least square regression [2].

Table 4 Soil quality prediction time measure of the proposed BQFS-TLDL method and other deep learning methods

Samples	Soil quality prediction time (ms)		
	BQFS-TLDL	Random forest ensemble	Partial least square regression
2800	1540	1960	2464
5600	1735	2145	2735
8400	2145	3125	3825
11200	2355	3355	4015
14000	2815	3895	4835
16800	3245	4215	5025
19600	3535	4495	5345
22400	4125	5135	5915
25200	4535	5845	6535
28000	5025	6035	7145

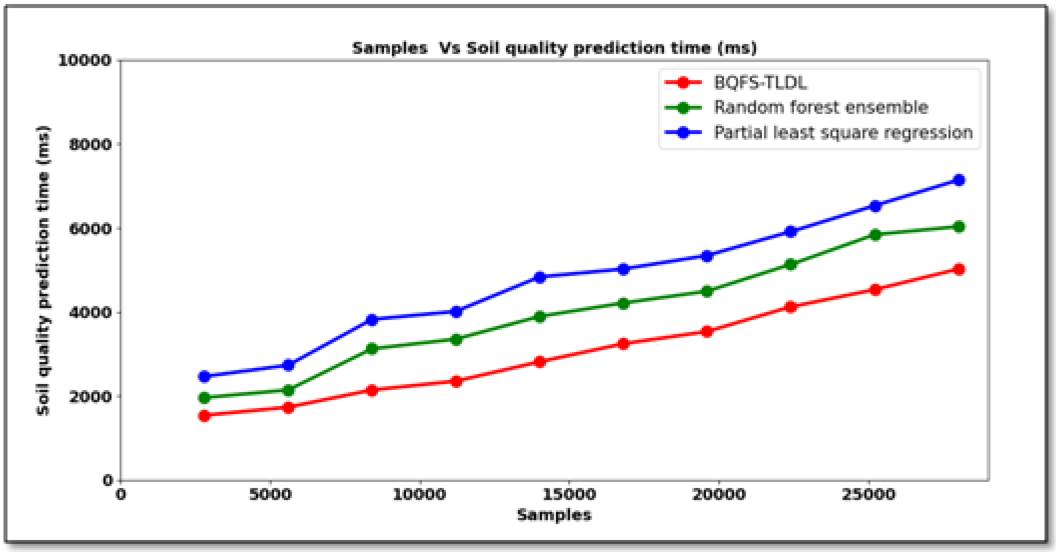

Figure 4 Soil quality prediction time results based on soil sample

Figure 4 illustrates soil quality prediction time results based on soil samples acquired from IoT soil moisture sensors set up in each of the fields with the IoT weather station positioned near the fields. With the seven different data transmitted via IoT devices, the prediction time involved was analyzed and plotted in the above figure. An increase in the graph is found in all the three methods. However, the prediction time was found to be decreasing using BQFS-TLDL upon comparison with Random forest ensemble [1] and Partial least square regression [2]. The decrease in time employing BQFS-TLDL method is used for Bipartite Quadratic Mutual Information and Fisher Score-based feature selection model. By applying this model, not only significant but precise features were obtained based on the bipartite function, considering mutual information and fisher score. This in turn assisted the BQFS-TLDL method in minimizing the soil quality prediction time by 23% and 36% compared to [1][2].

5.2 Case scenario 2: Soil quality prediction accuracy

In this section, soil quality prediction accuracy is evaluated. The second parameter of significance in predicting the soil quality is the rate of accuracy achieved in the process. In other words, accurate the prediction is made, significant soil features are said to be learnt, therefore, ensuring better yield to the farmers. The soil quality prediction accuracy is measured as given below.

$$P_{acc} = \sum_{i=1}^{n} \frac{S_{AP}}{S_i} * 100 \quad (17)$$

In (17), soil quality prediction accuracy 'P_{acc}' was calculated on samples 'S_i' involved in the simulation process according to three types of irrigation schedules (i.e., usual irrigation, less than crop needs and based on water loss) and the accurate prediction made 'S_{AP}' respectively. Table 5 given below lists the soil quality prediction accuracy for three different methods, BQFS-TLDL, Random forest ensemble [1] and Partial least square regression [2].

Table 5 Soil quality prediction accuracy measure of the proposed BQFS-TLDL method and other deep learning methods

Samples	Soil quality prediction accuracy (%)		
	BQFS-TLDL	Random forest ensemble	Partial least square regression
2800	97.32	95.89	93.75
5600	96.25	95.25	93.45
8400	96.05	94.35	93.15
11200	96	93.85	93
14000	95.35	93	92.25
16800	95.15	92.15	90
19600	95	92	89.35
22400	94.85	92	88.15
25200	94.25	91.35	86
28000	94	90	85.34

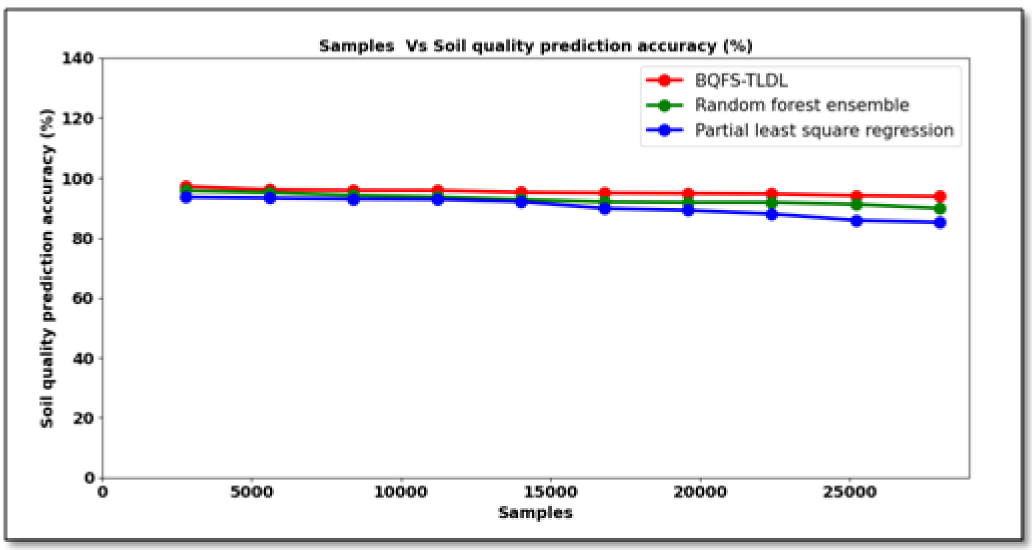

Figure 5 Soil quality prediction accuracy results based on soil sample

Figure 5 given above illustrates the graphical representation of soil quality prediction accuracy based on three different types of irrigation schedules, i.e., usual irrigation, less than crop needs and based on water loss respectively. From the above figure, a small deviation was observed in obtaining the accuracy rate though found to be in the decreasing trend when increasing the soil samples. This is owing to the reason that as soil sample sizes increases the prediction has to be made in 5-minute increments. Here, last four days were utilized for soil humidity in fields 1 and 3 and in 5-minute increments to be performed for the last six days pertaining to soil humidity in fields 2 and 4. However, the accuracy was found to be comparatively better using BQFS-TLDL upon comparison with Random forest ensemble [1] and Partial least square regression [2]. This is because of quadratic matrixto identify feature combinations that are not correlated with each other. Also with the aid of this quadratic matrix mathematically formulation is made for two different types of crops separately. This in turn assists in improving the soil quality prediction accuracy using BQFS-TLDL method as 3% and 6% compared to [1][2].

5.3 Case scenario 3: Precision

Precision is defined by the difference in measured values while performing repeated measurements of the same quantity. In other words, precision simply denotes the error that changes randomly each time during the repetition of measurement (i.e., soil quality prediction).

$$P = \frac{TP}{TP+FP} \quad (18)$$

In (18), precision value 'P' was calculated on true positive 'TP', (i.e., correct prediction of soil quality) as well as false positive 'FP' (such as falsely predicting results).Table 6 given below provides the results of precision acquired at different time intervals using the proposed method, BQFS-TLDL, Random forest ensemble [1] and Partial least square regression [2].

Table 6 Soil quality prediction accuracy measure of the proposed BQFS-TLDL method and other deep learning methods

Samples	Precision (%)		
	BQFS-TLDL	Random forest ensemble	Partial least square regression
2800	0.97	0.96	0.93
5600	0.96	0.94	0.92
8400	0.94	0.92	0.9
11200	0.93	0.9	0.88
14000	0.93	0.89	0.86
16800	0.92	0.88	0.85
19600	0.91	0.86	0.84
22400	0.9	0.85	0.83
25200	0.89	0.83	0.82
28000	0.88	0.82	0.81

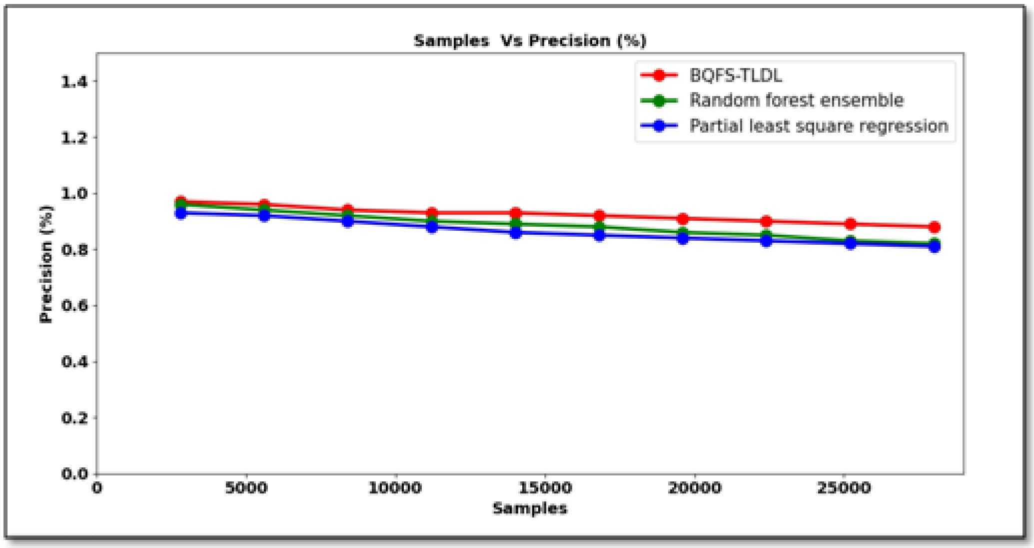

Figure 6 Precision results based on soil sample

Figure 6 given above shows the graphical representation of precision for 2800 soil samples obtain at different time intervals from four different fields. In above figure precision rate was comparatively higher using BQFS-TLDL upon comparison with Random forest ensemble [1] and Partial least square regression [2]. The increase in the precision rate using BQFS-TLDL method was false prediction results was lesser than [1] and [2]. The reason behind the minimization of false prediction results was due to the application of Triplet Loss-based Deep Learning for Soil Quality Prediction. By applying this model, Triplet Loss function was utilized in updating the weight. By applying this Triplet loss weight parameters were fine tuned that in turn made the interval between anchor and positive samples to be smaller than interval between anchor and negative samples. As a result, the precision rate was found to be comparatively better using BQFS-TLDL method by 4% in comparison with [1] and 7% in comparison with [2] respectively.

5.4 Case scenario 4: False positive rate

Finally, false positive rate 'FPR' was estimated on basis of amount of false positives 'FP' and amount of true negatives 'TN'. This is mathematically formulated as given below.

$$FPR = \frac{FP}{FP+TN} \tag{19}$$

Finally, table 7 given below provides the results of false positive rate obtained using the proposed method, BQFS-TLDL, Random forest ensemble [1] and Partial least square regression [2].

Table 7 False positive rate measure of the proposed BQFS-TLDL method and other deep learning methods

Samples	False positive rate (%)		
	BQFS-TLDL	Random forest ensemble	Partial least square regression
2800	0.02	0.03	0.04
5600	0.03	0.039	0.048
8400	0.035	0.045	0.05
11200	0.038	0.048	0.054
14000	0.04	0.051	0.059
16800	0.044	0.055	0.06
19600	0.046	0.057	0.062
22400	0.047	0.06	0.065
25200	0.049	0.062	0.07
28000	0.052	0.064	0.072

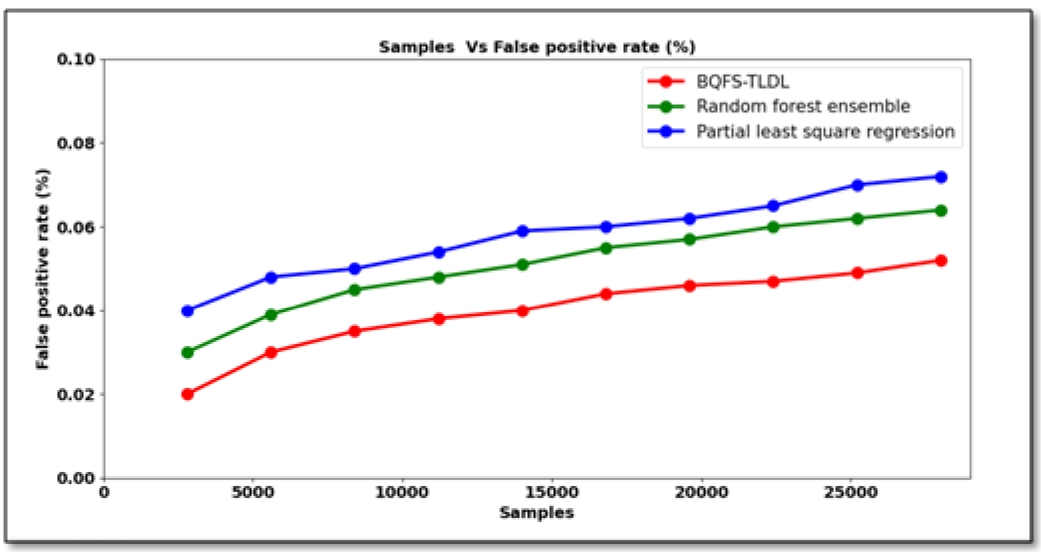

Figure 7 False positive rate results based on soil sample

Figure 7 illustrates false positive rate to various samples employing the thirteen different types of context data acquired from table 2 on a daily basis for maize and peanut separately. Also a linear trend is observed such as enhancing samples within false positive rate. This is due to the reason that by increasing the sample size, false positive rate reduces and true negative rate improves. However, false positive rate using BQFS-TLDL is found to be comparatively lesser with [1] and [2]. This is because of minimization of false positive rate was owing to the Triplet Loss-based Deep Learning algorithm for soil quality prediction. With this algorithm, for fine tuning the weight parameters makes the interval between anchor and positive samples to be smaller and the interval between anchor and negative samples to be larger. Also, by using long short term network, the relevant features and the information is combined in a weighted manner that in turn passes on to the next layer. Finally by comparison with expected output with model's output weights are updated. This in turn reduces the false positive rate using BQFS-TLDL as 22% and 31% compared to [1][2].

6. Conclusion

In this paper, the enhanced machine and deep learning model for IoT-based soil quality prediction is developed toassist in agriculture with IoT network and deep learning techniques. Here, with several context data in consideration and only significant features selected can in turn improve soil quality prediction in an accurate, precise and timely manner. The method also selected the significant feature based on the sample features employing Bipartite Quadratic by means of Mutual Information Coefficient (MIC) and Fisher Score (FS). Also, the proposed method is designed in such a manner to fine tune the weight from the LSTM by means of Triplet Loss function that minimizes false positive rate and increases accuracy. Evaluations are determined based on metrics namely soil quality prediction accuracy, soil quality prediction time, precision and false positive rates. BQFS-TLDL efficiently achieves better rate of accuracy with reduced false positive rates.

References

[1] IsayaKisekka, Srinivasa Rao Peddinti, William P. Kustas, Andrew J. McElrone, Nicolas Bambach‑Ortiz, Lynn McKee, Wim Bastiaanssen," Spatial–temporal modeling of root zone soil moisture dynamicsin a vineyard using machine learning and remote sensing", Irrigation Science, Springer, Jan 2022 [Random forest ensemble]

[2] Gilson Augusto Helfera, Jorge Luis Victória Barbosa, Ronaldo dos Santos,Adilson Ben da Costa, "A computational model for soil fertility prediction in ubiquitous agriculture", Computers and Electronics in Agriculture, Elsevier, Jun 2020 [Partial least square regression]

[3] StanisławGruszczynski, WojciechGruszczynski, "Supporting soil and land assessment with machine learning models using the Vis-NIR spectral response", Geoderma, Elsevier, Feb 2022

[4] Andrei Dornik, Marinela Adriana Chetan, Lucian Dragut¸ Daniel DorinDicu, Andrei Iliut, "Optimal scaling of predictors for digital mapping of soil properties", Geoderma, Elsevier, Sep 2021

[5] Nemanja Filipovi, SanjaBrdar, Gordan Mimi, Oskar Marko,Vladimir Crnojevi, "Regional soil moisture prediction system based onLong Short-Term Memory network", Biosystems Engineering, Elsevier, Dec 2021

[6] Qingliang Li, Yuheng Zhu, Wei Shangguan, Xuezhi Wang, Lu Li, Fanhua Yu, "An attention-

aware LSTM model for soil moisture and soil temperature prediction", Geoderma, Elsevier, Dec 2021

[7] G.Murugesan, Dr. B.Radha, "Soil Data Classification Using Attribute Group Rank with Filter Based Instance Selection Model", International Journal of Scientific &Technology Research Volume 9, Issue 06, June 2020

[8] Freddy A. Diaz-Gonzalez, Jose Vuelvas, Carlos A. Correa, Victoria E. Vallejo, D. Patino, "Machine learning and remote sensing techniques applied to estimate soil indicators – Review", Ecological Indicators, Elsevier, Dec 2021

[9] Thomas van Klompenburg, AyalewKassahun, CagatayCatal, "Crop yield prediction using machine learning: A systematic literature review", Computers and Electronics in Agriculture, Elsevier, Aug 2020

[10] MamunurRashid, BiftaSamaBari, YusriYusup, Mohamad AnuarKamaduddin, NuzhatKhan, "A Comprehensive Review of Crop YieldPrediction Using Machine Learning ApproachesWith Special Emphasis on PalmOil Yield Prediction", IEEE Access, Apr 2021

[11] Zeynep Ünal, "Smart Farming Becomes Even Smarter WithDeep Learning_A Bibliographical Analysis", IEEE Access, Jun3 2020

[12] Abhinav Sharma, Arpit Jain, Prateek Gupta, Vinay Chowdary, "Machine Learning Applications for PrecisionAgriculture: A Comprehensive Review", IEEE Access, Dec 2020

[13] Machbah Uddin, Md. Rakib Hassan, "A novel feature based algorithm for soil type classification", Springer, Feb 2022

[14] M.S. Suchithra, Maya L. Pai, "Improving the prediction accuracy of soil nutrientclassification by optimizing extreme learningmachine parameters", Information Processing in Agriculture, Elsevier, May 2019

[15] DhivyaElavarasan and P. M. DurairajVincent, "Crop Yield Prediction Using Deep ReinforcementLearning Model for SustainableAgrarian Applications", IEEE Access, Apr 2020

[16] Paul Simfukwe, Paul W. Hill, Bridget A. Emmet3, Davey L. Jones, "Identification and predictability of soil qualityindicators from conventional soil andvegetation classifications", PLOS ONE | https://doi.org/10.1371/journal.pone.0248665 October 22, 2021

[17] Paul Simfukwe, Paul W. Hill, Bridget A. Emmett, Davey L. Jones, "Identification and predictability of soil qualityindicators from conventional soil andvegetation classifications", PLOS ONE | https://doi.org/10.1371/journal.pone.0248665 October 22, 2021

[18] Naser Shiri, Jalal Shiri, Zaher Mundher Yaseen, Sungwon Kim, Il-Moon Chung, Vahid Nourani, Mohammad Zounemat-Kermani, "Development of artificial intelligence modelsfor well groundwater quality simulation:Different modeling scenarios", PLOS ONE | https://doi.org/10.1371/journal.pone.0251510 May 27, 2021

[19] Johnathon Shook, TryambakGangopadhyay, Linjiang Wu,Baskar Ganapathysubramanian, Soumik Sarkar, Asheesh K. Singh, "Crop yield prediction integrating genotypeand weather variables using deep learning", PLOS ONE | https://doi.org/10.1371/journal.pone.0252402 June 17, 2021

[20] Yu Cai, Wengang Zheng, Xin Zhang, Lili Zhangzhong, XuzhangXue, "Research on soil moisture prediction modelbased on deep learning", PLOS ONE | https://doi.org/10.1371/journal.pone.0214508 April 3, 2019