Development of pedo-transfer functions for estimating soil erodibility and soil aggregate stability from basic soil properties in submontane areas using statistical methods and machine learning techniques for Northern India land

Abstract: Accelerated soil erosion in the lower *Shiwalik hills in Submontane area of Northern India land* results in the deterioration of soil physical quality. Soil aggregation and soil erodibility are important parameters indicating soil physical quality but the quantification of these parameters is complex so efforts have been done to estimate these properties from easily measurable soil characteristics by using pedo-transfer functions (PTFs). Statistical PTFs are available for estimating of these properties but they don't explain the sufficient variability in data. Machine learning techniques may play an important role in this context. Therefore, the present study was planned to compare PTFs developed using statistical and machine learning techniques for the *kandi* region of Punjab. The basic soil physico-chemical properties were measured across four locations with five land uses at each location at three depths and with three replications. Three data sets were prepared for these soil properties. When dataset 1, having six basic soil properties, was used for the estimation of mean weight diameter (MWD) and erodibility (K), the prediction using an artificial neural network (ANN) was slightly better than the generalized linear model (GLM). When dataset 2, having those six basic soil properties which were having a high correlation with soil structural parameters, was used for the estimation of MWD and K, the prediction using GLM was slightly better than ANN. When dataset 3, having all 11 basic soil properties, was used for the estimation of MWD and K, the prediction using ANN was significantly better than GLM. So, it may be concluded that ANN performs better for a large set of data and ~~for~~ a complex system having a greater number of variables whereas for a small set of data and for a simple system having ~~less~~ fewer variables, the statistical methods perform better.

Keywords: erodibility; aggregate stability; PTFs; machine learning; artificial neural network; generalized linear model

1. Introduction

Land degradation is a major issue ~~which~~ that affects the capability of ecosystem services provided by the soil. The decline in soil quality caused by anthropogenic activities has been a global issue during the previous century and still, it has remained high on the international agenda during the current century because of its impact on world food security and ~~the~~ environmental quality. The lower Shiwalik hills in Submontane area of Northern India ~~is~~ are suffering from severe soil erosion resulting in the deterioration of soil physical quality in the region [1]. ~~Stability~~ The stability of soil aggregates is considered as one of the most important indicators of soil physical quality. It is the measure of the resistance of soil aggregates against ~~the~~ structural decomposition because of raindrop impact, running water, or wind [2-3]. Aggregate stability is a soil characteristic ~~which~~ that is often linked to soil erodibility [4]. Soil cementing agents like clay, silt, and organic matter which result in aggregate stability are usually correlated with soil aggregate stability [5]. The soil erodibility which is the measure of the resistance offered by the soil to both detachment and transport processes of soil erosion, is an inherent property of the soil. It is influenced by soil's physical characteristics including texture, structure, organic matter, and chemical characteristics. Assessment of soil erodibility is important for erosion prediction and for planning suitable soil conservation measures. Mean Weight Diameter (MWD), Geometric Mean Weight Diameter (GMD), and percentage of Water Stable Aggregates (WSA) are the common parameters representing soil aggregate stability [6-7]. However, out of these indices, the MWD is the most widely used indicator for quantification of soil aggregate stability [8]. For the measurement of ~~the~~ aggregate stability, the most common method is the wet sieving method [9]. For the measurement of aggregate stability, Le bissonnais [7] proposed a standard wet sieving method ~~which~~ that consists of three treatments. These include fast wetting leading to slaking, slow wetting leading to microcracking, and stirring of pre wetted aggregates for mechanical breakdown. However, evaluation of soil aggregate stability using these methods is time consuming and expensive. So, to overcome this difficulty, Pedo-transfer functions (PTFs) have been developed for predicting aggregate stability [10-12]. For example, easy ~~to~~ -to-measure soil parameters like organic carbon, particle size distribution, and bulk density are used in empirical multilinear regression-based models for the estimation of complex soil properties like mean weight diameter [13]. Researchers have developed PTFs to estimate soil erodibility also from basic soil properties under various conditions. These PTFs have been used as ~~an~~ input for environmental simulation models. The pedo-transfer functions are used basically to translate the raw soil data into ~~the~~ more useful information. These PTFs include linear, logarithmic, and other statistical models using various basic soil properties for the estimation of soil aggregate stability and soil erodibility. There is normally poor performance of the ~~regression~~ regression-based PTFs as they require prior information about input–output relationships. The statistical regression models require prior information about the relationship between independent and dependent soil properties and on the other hand for the neural network model there is no need ~~of~~ for this type of prior information. Recently artificial intelligence in the form of machine learning techniques is also being employed in predictive models. Machine learning is the combination of processes that gives machines ~~an~~ the ability to learn without the use of specific software program~~me~~s. Machine learning methods like K Nearest Neighbor (KNN), Cubist, Artificial neural network (ANN), and Random Forest (RF) approaches have been deployed recently in the development of PTFs. Out of these approaches, ANN is a simplified model representing the structure of the biological neural network in which interconnected processing units are organized in a specific topology. Multiple layers of information are arranged using ~~a~~ several ~~number of~~ nodes. These nodes include an input layer for feeding the data into the system, one or more intermediate hidden layers in which the learning takes place, and an output layer for providing the decision or prediction. No prior relationship between the input and output variables is required for machine learning techniques and it is one of their major advantage [14-15]. Although several statistical PTFs are available for estimating soil aggregate stability and soil erodibility from basic soil properties, still their standardization for identifying minimum data set is required for *kandi* region of Punjab. Machine learning techniques may play an important role in this context. Therefore, the present study has been planned to compare existing PTFs with developed PTFs using statistical and machine learning techniques with the objectives of the Development of pedo-transfer functions for estimating soil erodibility and soil aggregate stability from basic soil properties using statistical methods and machine learning techniques. Comparison of estimated soil erodibility and soil aggregate stability using PTFs developed statistically, PTFs developed through machine learning techniques. Better estimation of soil erodibility and soil aggregate stability from easily measurable soil properties using PTFs may lead to better estimation of soil erosion which may help in the management of soil erosion.

2. Materials and Methods

Soil samples and Soil properties

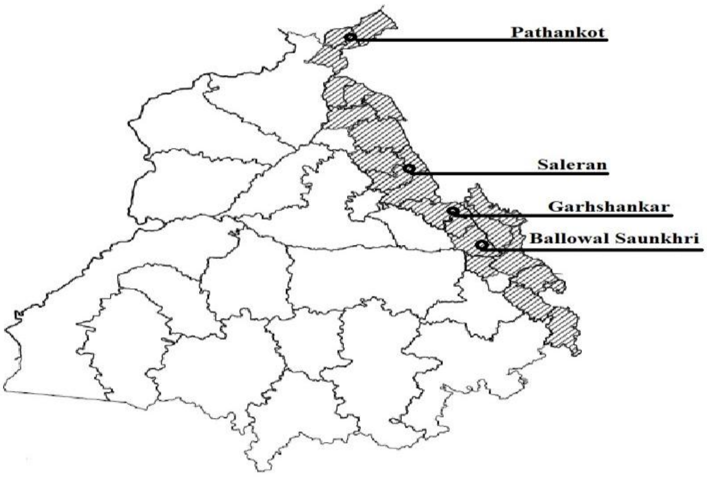


Figure 1. Location of the study area

The study was conducted at four locations in submontanous kandi region of Punjab in the districts of Pathankot (32^{0}33`N, 75$^{0}$69`E), Saleran (31^{0}59`N, 75$^{0}$97`E), Garhshankar (31^{0}28`N, 76$^{0}$21`E) and Ballowal Saunkhri (31^{0}09`N, 76$^{0}$38`E). The *Kandi* region's climate varied from semi-arid to ~~sub-~~ sub-humid. The yearly rainfall in the area is around 1090 ± 340 mm. The rainfall distribution is bimodal, with 75–80% of total rainfall falling between June and September and 20–25% falling between the winter months (October to March). Soil samples were taken from Agroforestry, Grassland, Horticulture, Forestry, and Agriculture in each of the four locations. Soils were sampled at three depths within each land use: 0-7.5, 7.5-15, and 15-30 cm. A total of 180 data points ~~was~~ were there by taking 4 locations, 3 replications, 5 land uses, and three depths from each location. Soil samples were analysed for basic soil properties like pH, EC, OC, CEC, calcium carbonate, bulk density, Fe, and soil particle size analysis and applied soil properties like soil aggregation and soil erodibility.

Figure 2. Some photographs of selected land uses

Soil sample analysis

Soil samples were ~~air~~ air-dried, crushed, and sieved, using a 2mm sieve before being analysed for a variety of physico-chemical properties. Undisturbed soil samples were also taken in the form of huge clods of roughly 40-50 cm diameter using a spade from 0-15 and 15-30 cm depths at four locations in each land use. The clods were carefully transported to the laboratory and dropped from a height of 90-100 cm on grassy ground, breaking at natural weak spots. Wet sieving was done with the resultant aggregates. Using cores, separate samples were taken for bulk density assessments. Soil texture was analyzed by International pipette method [16], Organic carbon by Rapid titration method [17], Calcium Carbonate by Puri's method [18], Cation exchange capacity by Ammonium acetate extraction method [19], pH by 1:2 soil water suspension [20], Electrical conductivity [21], Aggregate stability by Wet sieving method using Yoder apparatus [22], Bulk density by Core method [23], Iron by Atomic absorption spectroscopy [24]. The nomographic expression proposed by [25] can be used to estimate K from easily observable soil parameters such as texture, organic content, structure, and permeability. Singh and Khera [1] provided a modified technique for estimating K.

$$K = M^{1.14}(10^{-7})\ (12-\alpha) + 4.28(10^{-3})\ (\beta-2) + 3.29(10^{-3})\ (\gamma-3)$$

Where~~,~~ M =M was calculated as 100 X (percentage of aggregates and primary particles<2.0 mm). α =Organic matter (%) β = structure code γ = permeability rating.

Machine learning technique

An open-source **Big ML** software was used to estimate the soil aggregate stability and soil erodibility for machine learning techniques and Multi-linear regression equation (GLM). For machine learning and Generalized linear model, training and testing of data was done. For training, 70% data was used and the remaining 30% data was used for testing.

~~Total~~ A total of 180 data points (Four locations X five land uses X three depths X three replications) were generated for 11 basic soil characteristics and three applied soil properties. Three data sets were prepared for these soil properties as described below:

Dataset 1 Properties like Sand, silt, clay, bulk density, EC, and Organic carbon were used in data set 1. These are the properties ~~which~~ that are available commonly in the literature also. This data set was prepared both from research data as well as from secondary literature.

Dataset 2 Properties which show a significant correlation in the correlation matrix for (MWD and K) were used in ~~dataset~~ Dataset 2. These are clay, Fe, calcium carbonate, pH, EC, OC, and BD.

Dataset 3 (K, MWD) All 11 properties from research data were used in dataset 3. These are coarse sand, fine sand, silt, clay, Fe, calcium carbonate, pH, EC, OC, BD, and Cation exchange capacity).

Secondary Data Obtain from Literature for MWD and K

Table Ⅰ. Secondary data obtained from the literature for MWD

References	PTFs (Basic properties)	Applied property
[26]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[27]	Clay, EC, Sand, OC, Silt, Bulk density	MWD

[28]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[29]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[30]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[31]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[32]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[33]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[34]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[35]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[36]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[37]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[38]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[39]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[40]	Clay, EC, Sand, OC, Silt, Bulk density	MWD
[41]	Clay, EC, Sand, OC, Silt, Bulk density	MWD

Table Ⅱ. Secondary data obtained from the literature for K

References	PTFs (Basic properties)	Applied property
[42]	Clay, EC, Sand, OC, Silt, Bulk density	K
[43]	Clay, EC, Sand, OC, Silt, Bulk density	K
[44]	Clay, EC, Sand, OC, Silt, Bulk density	K
[45]	Clay, EC, Sand, OC, Silt, Bulk density	K
[46]	Clay, EC, Sand, OC, Silt, Bulk density	K
[47]	Clay, EC, Sand, OC, Silt, Bulk density	K
[29]	Clay, EC, Sand, OC, Silt, Bulk density	K
[48]	Clay, EC, Sand, OC, Silt, Bulk density	K

Evaluation of PTFs

Different regression metrics were used to evaluate the model

Root mean square error

~~Lower~~ The lower the RMSE value, the better the model performance. RMSE was used while calibration of the model ~~for the purpose of finding~~to find the most sensitive parameters. This is a measure of the model's real inaccuracy and its calculated as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)}{n}}$$

Coefficient of correlation

The correlation ~~Coefficient~~ coefficient indicates how close the observed and projected regression lines are to an ideal match. This coefficient is normally between -1 to +1 and is estimated ~~it~~ as follows:

$$R = \frac{\sum_{i=1}^{n}(o_i - o_{avg})\sum_{i=1}^{n}(s_i - s_{avg})}{\sqrt{\sum_{i=1}^{n}(O_i - O_{avg})^2 \sum_{i=1}^{n}(s_i - s_{avg})^2}}$$

Coefficient of determination (R^2)

$$R^2 = 1 - \frac{(\sum_{i=1}^{n}(P_i - O)^2)/(n-k)}{(\sum_{i=1}^{n}(O_i - O)^2)/(n-1)}$$

MAE (Mean Absolute Error)

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i|$$

Where O_i is the observed aggregate stability and P_i is the anticipated aggregate stability, respectively, O is the mean of the observed values, k is the total number of ~~explanatories~~explanatory, and n is the number of values.

3. Results

Basic Soil properties

The Basic soil properties were analysed and results were concluded, the pH varies from 6.6-7.7, EC Varies from 0.10-0.23dS m^{-1}, Organic Carbon is 0.59-0.92%, Cation Exchange Capacity: is 8.41-13.71 C mol kg^{-1}, Calcium Carbonate 0.06-0.09%, Bulk density: 1.27-1.48 Mg m^{-3}, Fe content varies from 11.9-21.9 mg kg^{-1}, Mean weight diameter 0.46-2.59 mm, K erodibility factor varies from 0.16-0.33. Textural class at Pathankot and Saleran was loamy sand, at Ballowal Saunkhri it was sandy loam, and at Garhshankar it was sandy clay loam.

Development of PTFs statistically for aggregate stability and soil erodibility

Estimating MWD

Using Dataset 1 soil properties

MWD**=** 4.21+0.07*Sand + 0.02*Silt + 0.01*Clay + 0.87***OC – 2.36*BD – 11.23 *EC**

R^2=0.61, MAE =0.46, MSE =0.35, RMSE =0.59

Using Dataset 2 soil properties

MWD = 5.23-0.009*pH-**11.43*EC**+ **1.10*OC** + **0.53*$CaCO_3$**- **2.10*BD** + 0.005*Clay+0.009*Fe

R^2=0.56, MAE =0.38, MSE =0.56, RMSE=0.74

Using Dataset 3 soil properties

MWD = 4.71+0.03*pH-**9.15*EC** + **0.87*OC**+ 0.001*CEC + **0.37*$CaCO_3$**- **2.61*BD** + 0.01*Coarse sand-0.02*Fine sand + 0.03*Silt -0.01*Clay + **0.02*Fe**

R^2=0.59, MAE =0.50, MSE =0.44, RMSE=0.66

Estimating soil erodibility

Using Dataset 1 soil properties

K **=** -0.76+ 0.005*Sand + **0.01*Silt**- **0.01*Clay**- **0.02*OC**+ **0.20*BD**+ **0.93*EC**

R^2=0.65, MAE =0.04, MSE= 0, RMSE=0

Using Dataset 2 soil properties

K = 0.03 +0.06*pH +**0.82*EC – 0.005*OC +** 0.07*$CaCO_3$ **+ 0.10***BD - **0.05*Clay-0.04*Fe**

R^2=0.85, MAE =0.03, MSE=0, RMSE=0

Using Dataset 3 soil properties

K = -0.40+0.04*pH +**0.63*EC – 0.001*OC** + 0.008*CEC + 0.076*$CaCO_3$+ **0.16*BD** + 0.03* Coarse sand + 0.003* Fine sand + 0.006* Silt -**0.003*** **Clay** -**0.007*Fe**

R^2=0.73, MAE =0.04, MSE =0, RMSE=0

Development of PTFs by machine learning for soil erodibility and aggregate stability

Estimating MWD

Using Dataset 1 soil properties

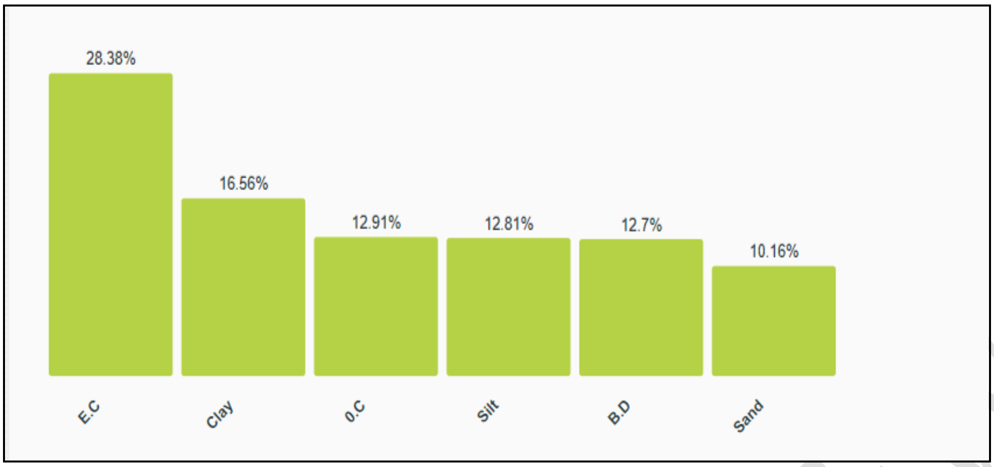
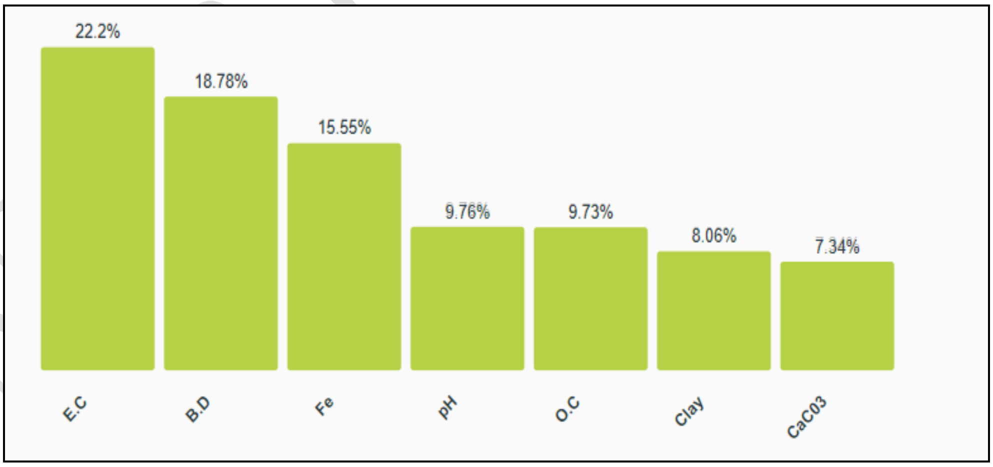


Figure 3. Properties of field importance and Different Regression metrics for MWD using ~~dataset~~ Dataset 1

Using Dataset 2 soil properties



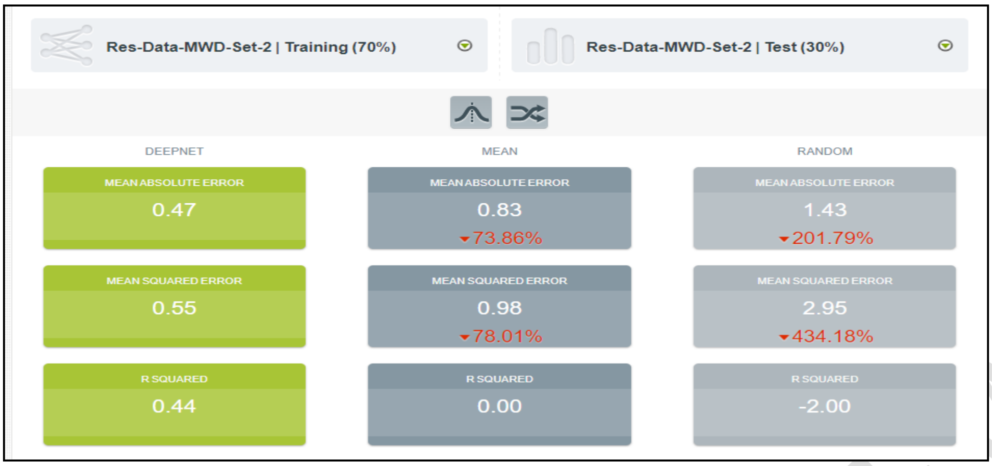


Figure 4. Properties of field importance and Different Regression metrics for MWD using ~~dataset~~ Dataset 2

Using Dataset 3 soil properties

Figure 5. Properties of field importance and Different Regression metrics for MWD using ~~dataset~~ Dataset 3

Estimating soil erodibility

Using Dataset 1 soil properties

Figure 6. Properties of field importance and Different Regression metrics for K using dataset 1

Using Dataset 2 soil properties

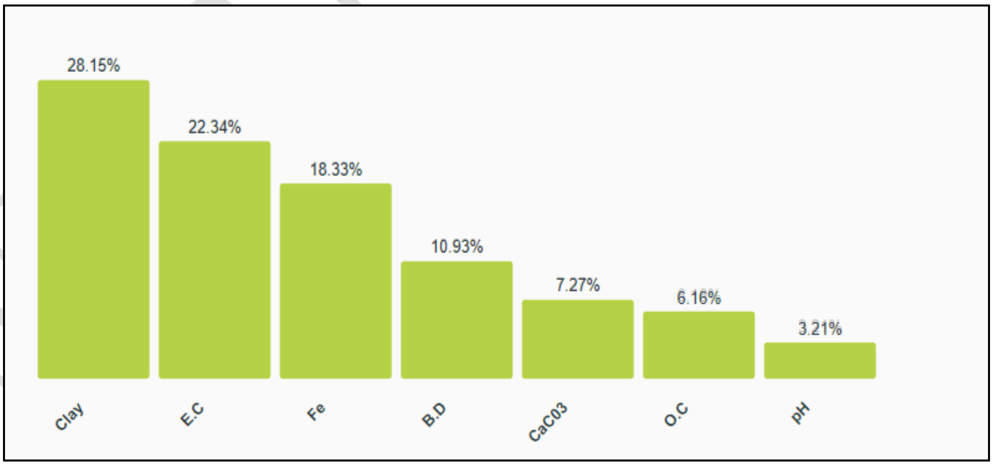
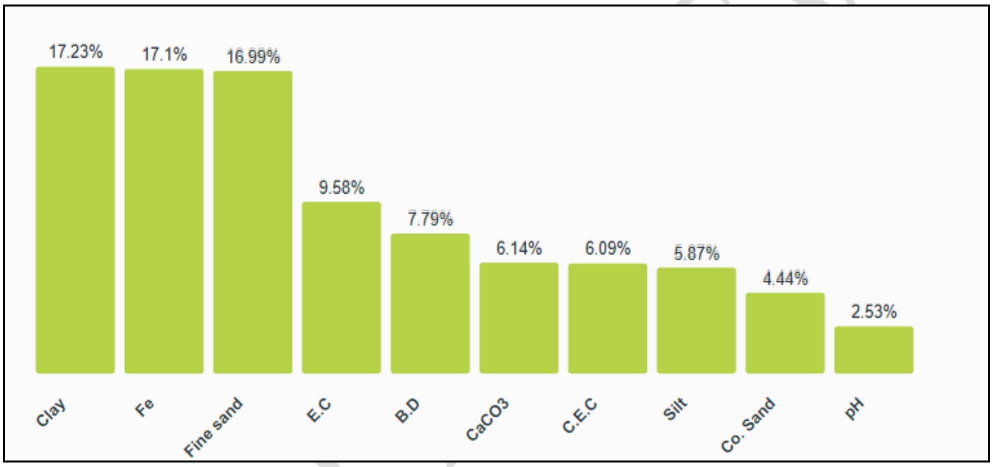


Figure 7. Properties of field importance and Different Regression metrics for K using dataset 2

Using Dataset 3 soil properties



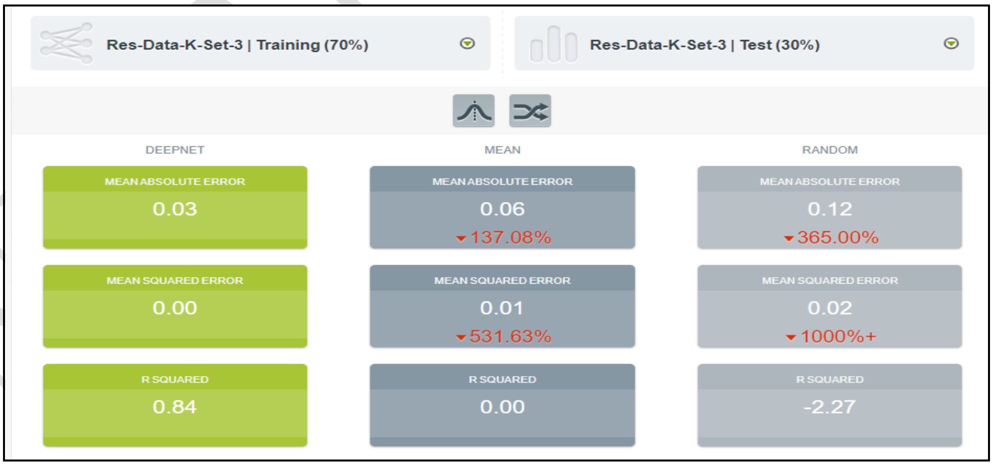


Figure 8. Properties of field importance and Different Regression metrics for K using dataset 3

Comparison between different PTFs developed through machine learning and statistically

The value of comparing the ANN and GLM models, the results showed that ANN explained the variability much better than the GLM for dataset 3, in which eleven soil properties were used, for the prediction of all three complex soil properties i.e. K, MWD. This is also evident from the values of MAE, MSE, and RMSE obtained using GLM and ANN. Whereas, for data set 1 and data set 2, where the number of basic soil properties used was less, the results were not consistent. Comparing the ANN and GLM it was concluded that ANN performs better for a large set of data and ~~for~~ a complex system having a greater number of variables whereas for a small set of data and for a simple system having ~~less~~ fewer variables the statistical methods perform better.

Table 3: Comparing ~~of~~ ANN and GLM

Different Data set	**MAE**		**MSE**		**RMSE**		**R^2**	
	GLM	**ANN**	**GLM**	**ANN**	**GLM**	**ANN**	**GLM**	**ANN**
Data set 1 (MWD)	0.46	0.46	0.35	0.37	0.59	0.6	0.61	0.64
Data set 1 (K)	0.03	0.03	0.00	0.00	0.00	0.00	0.65	0.79
Data set 2 (MWD)	0.38	0.47	0.56	0.55	0.74	0.74	0.56	0.44
Data set 2 (K)	0.03	0.03	0.00	0.00	0.00	0.00	0.85	0.73
Data set 3 (MWD)	0.50	0.22	0.44	0.10	0.66	0.31	0.59	0.88
Data set 3 (K)	0.03	0.03	0.00	0.00	0.00	0.00	0.73	0.84

5. Conclusions

Machine learning (ANN) and Statistical model (multi-linear regression / GLM) was used for developing PTFs for aggregate stability and soil erodibility. Three types of datasets were made for basic soil properties and were used for prediction of MWD and K. 70% of the total available, research and literature data was utilized to train the model, while 30% was used to test the model. For dataset 1, using the GLM model, the R^2 values between actual and predicted MWD and K were 0.61 and 0.65, respectively. Whereas for the same dataset 1, using the ANN model, the R^2 values between actual and predicted MWD and K were 0.64 and 0.79, respectively. For dataset 2, using the GLM model, the R^2 values between actual and predicted MWD and K were 0.56 and 0.85, respectively. Whereas for the same dataset 2, using the ANN model, the R^2 values between actual and predicted MWD and K were 0.44 and 0.73, respectively. For dataset 3, using the GLM model, the R^2 values between actual and predicted MWD and K were 0.73 and 0.59, respectively. Whereas for the same dataset 3, using the ANN model, the R^2 values between actual and predicted MWD and K were 0.88 and 0.84, respectively. So, it may be concluded that ANN performs better for a large set of data and

~~for~~ a complex system having a greater number of variables whereas for a small set of data and for a simple system having ~~less~~ fewer variables, the statistical methods perform better.

References

[1] Singh MJ; Khera KL. Physical indicators of soil quality in relation to soil erodibility under different land uses. *Arid Land Research and Management* **2009**, 23, 152-167.

[2] Canasveras JC, Barron V, Del Campillo MC, Torrent J; Gomez JA. Estimation of aggregate stability indices in Mediterranean soils by diffuse reflectance spectroscopy. *Geoderma* **2010;** 158, 78-84.

[3] Besalatpour AA, Ayoubi S, Hajabbasi MA, Mosaddeghi MR and Schulin R. Estimating wet soil aggregate stability from easily available properties in a highly mountainous watershed. *Catena* **2013;** 111, 72-79.

[4] Kukal SS, Sur HS and Gill SS. Factors responsible for soil erosion hazard in Submontane Punjab, India. *Soil Use and Management* **1991;** 7, 38-44.

[5] Tejada M and González JL. The relationship between erodibility and erosion in ~~a~~ soil treated with two amendments. *Soil Tillage Research* **2006;** 9, 186–198.

[6] Calero N, Barron V and Torrent J. Water dispersable clay inn calcareous soils of southwestern Spain. *Catena* **2008;** 74. 22-30.

[7] Le-Bissonnais Y. Aggregate stability and assessment of soil crustability and erodibility: I theory and methodology. *European Journal of Soil Science* **1996;** 47, 1-425-137.

[8] Choudhary SG, Srivastava S, Singh R, Chaudhari SK, Sharma DK, Singh SK and Sarkar D. Tillage and residue management effects on soil aggregation, organic carbon dynamics, and yield attribute in rice-wheat cropping system under reclaimed sodic soil. *Soil Tillage Res*earch **2014;** 136, 76-83.

[9] Le-Bissonnais Y. Aggregate stability and assessment of soil crustability and erodibility: I theory and methodology. *European Journal of Soil Science* **2016;** 67, 1-21.

[10] Minasny B and Mcbratney AB. The neuro-m methods for fitting neural network parametric pedotransfer functions. *Soil Science Society of America Journal* **2002;** 66, 352-361.

[11] Minasny BJ, Hopmans JW, Harter T, Eching SO, Tuli A, and Denton MA. Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. *Soil Science Society of America Journal* **2004;** 68, 417-429.

[12] Fatichi SD, Walko R, Vereecken H, Young MH, Ghezzehei TA, Hengl T, Kollet S, Agam N and Avisaar R. Soil structure is an important omission in earth system models. *Nature Communication* **2020.**

[13] Bhattacharya P, Maity PP, Ray M and Krishnan P. Comparison of artificial neural network and multi-linear regression for prediction of field capacity soil moisture content. *Journal of Agricultural Physics* **2018;** 18(2), 173-180.

[14] Gocic M, Petkovic D, Shamshirband S and Kamsin A. Comparative analysis of reference evapotranspiration equations modelling by extreme learning machine. *Computers and Electronics in Agriculture* **2016;** 127(6), 56-63.

[15] Jafarzadeh AA, Pal M, Servati M, Fazelifard MH and Ghorbani MA. Comparative analysis of support vector machine and artificial neural network models for soil cation exchange capacity prediction. *Int J Env Sci Technol* **2016;** 13(1), 87-96.

[16] Piper C S. Soil and Plant Analysis. Bombay: Hans Publisher (1966).

[17] Walkley, A. and I. A. Black. An examination of Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. Soil Sci. **1934;** 37, 29-37.

[18] Puri, A.N. A new method for estimating total carbonates in soils. Bull Imp Agri Res **1930;** 206, 7.

[19 Sumner, M.E., Miller, W.P., Cations exchange capacity and Exchange Coefficients, In: Sparks DL, (ed.). Methods of soil analysis, Part 3- chemical methods. Agronomy Monograph **1996**; 9, 1201-1230.

[20] Thomas, G. W. Soil pH and soil acidity. Methods of soil analysis: part 3 chemical methods **1996;** 5, 475-490.

[21] Rhoades, J. D. Salinity: Electrical conductivity and total dissolved solids. *Methods of soil analysis: Part 3 Chemical methods* **1996;** *5*, 417-435.

[22] Yoder, R.E. "A direct method of aggregate analysis of soils and a study of the physical nature of erosion losses", Journal of American Society of Agriculture **1936;** 28, 337-351.

[23] Blake, G. R. Bulk density. *Methods of Soil Analysis: Part 1 Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling* **1965;** *9*, 374-390.

[24] Lindsay, W. L., & Norvell, W. Development of a DTPA soil test for zinc, iron, manganese, and copper. *Soil science society of America journal* **1978;** *42*(3), 421-428.

[25] Wischmeier WH, Johnson CB, Cross BV A soil erodibility nomograph for farmland and construction sites. Journal of Soil and Water Conservation **1971;** 26, 189-92.

[26] Bhattacharya P, Mity P P, Ray Mrinmoy and Mridha N Prediction of mean weight diameter of soil using machine learning approaches. IARI New Delhi **2021**.

[27] Yaseen, Z. M. An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: Review, challenges and solutions. *Chemosphere* **2021;** *277*, 130126.

[28] Dhaliwal S S, Singh B, Sharma B D and Khera K L Soil quality and yield trends of different crops in low productive submontane tract and highly productive area in Punjab, India. *Indian J Dryland Agric Res Develop* **2009;** 24:39-45.

[29] Singh, K. B., Jalota, S. K., & Sharma, B. D. Effect of continuous rice–wheat rotation on soil properties from four agro-ecosystems of Indian Punjab. *Communications in soil science and plant analysis* **2009;** *40*(17-18), 2945-2958.

[30] Kukal S S, Manmeet-Kaur, Bawa S S and Gupta N Water-drop stability of natural soil aggregates from different land uses after treatment with polyvinyl alcohol. *Catena* **2007;** 70, 475-79.

[31] Igwe C A and Nwokocha D Influence of soil properties on the aggregate stability of a highly degraded tropical soil in eastern Nigeria. *Int Agrophys* **2005;** 19, 131-139

[32] Saygin S D, Erpul G and Basarn M Comparsion of aggregate stability measurement methods for clay rich soils in asartepe catchment of turkey. *Land Degrad Dev* **2017;** 28, 199-206.

[33] Ghosh PK, Saha R, Gupta JJ, Ramesh T, Das A, Lama TD, Munda GC, Bordoloi JS, Verma MR, Ngachan SV. Long-term effect of pastures on soil quality in acid soil of North-East India. Soil Research. Jun 30 **2009;** 47(4), 372-9.

[34] Dutta M, Diengdoh J and Ram S Physico-chemical properties of west khasi hills soils of Meghalaya in relation to land uses. *Asian J Soil Sci* **2015; 10**, 288-94.

[35] Haghighi F, Gorji M and Shorafa M A study of the effects of land use changes on soil physical properties and organic matter. *Land Degrad Develop* **2010;** 21, 496-502.

[36] Canasveras J C, Barron V, Del Campillo M C, Torrent J and Gomez J A Estimation of aggregate stability indices in Mediterranean soils by diffuse reflectance spectroscopy. *Geoderma* **2010; 158**, 78-84.

[37] Hati K M, Swarup A, Dwivedi A K, Mishra A K and Bandyopadhyay K K Changes in soil physical properties and organic carbon status at the topsoil horizon of the Vertisol of central India after 28 years of continuous cropping, fertilization and manuring. *Agric Ecosyst Environ* **2007;** 119, 127-34

[38] Golchin A and Asgari H Land use effects on soil quality indicators in north-eastern Iran. *Aust J Soil Res* **2008;** 46, 27-36.

[39] Dutta N, Dutta S, Karmakar RM. Soil aggregation and erodibility indices under different land uses in Jorhat district of Assam. Journal of Soil and Water Conservation. **2016**; 15(4), 284-91.

[40] Kalhoro SA, Xu X, Chen W, Hua R, Raza S, Ding K. Effects of different land-use systems on soil aggregates: a case study of the Loess Plateau (Northern China). Sustainability. Aug 1 **2017;** 9(8), 1349.

[41] Somasundaram J, Singh R K, Parandiya A K, Ali S, Chauhan V, Nishant K. Sinha, Lakaria B L, Saha R, Chaudhary R S, Coumar M V, Singh R K and Simaiya R R Soil Properties under Different Land Use Systems in Parts of Chambal Region of Rajasthan. *J Agric Phys* **2013;** 13, 139-47.

[42] Olaniya M, Bora P K, Das S and Chanu P H Soil erodibility indices under different land uses in district of Meghalaya. *Nature* **2020;** 10, 149-86

[43] Belasri A, Lakhouili A and Halima O I Soil erodibility mapping and its correlation with soil properties of Oued EI makhazine watershed, Morocoo. *J Mater Environ Sci* **2017;** 8, 3208-15

[44] Chandel S, Hadda M S and Mahal A K Soil quality assessment through minimum data set under different land uses of sub montane Punjab. *Commun Soil Sci Plant Anal* **2018;** 1532-2416.

[45] Chukwuocha N, Amangabara G T and Amaechi C Determination of the erodibility status of some soils in IKEDURU local government area of IMO state, Nigeria. *Int J Geo Earth Env Sci* **2014;** 4, 236-43.

[46] Wang H, Zhang G, Li N, Zhang B and Yang H variation in soil erodibility under five typical land uses in a small watershed on the loess plateau china. *Catena* **2019;** 174, 24-35

[47] Stanchi S, Falsone G and Bonifacio E Soil aggregation, erodibility and erosion rates in mountain soils. *Solid earth* **2015;** 6, 403-414.

[48] Olubanjo O O Evaluation of soil erodibility under different land uses. The federal university of technology, Akure Nigeria **2017**.