Prediction Of Soil Texture Using Convolution Neural Network with Enhanced Regression Model

K. Anandan[1*], R. Shankar[2], S. Duraisamy[3]

Abstract

The chemical properties in the soil are most important factors to predict the texture of soil and it helps to decide about the farming accordingly. There are several chemical properties are present in different region and only few taken for prediction of texture in the proposed research. The proposed research has taken only eight parameters from the soil to predict the texture such as organic carbon, cation exchange capacity, nitrogen, pH level, clay, sand, potassium and phosphorous. By implementing these parameters with convolution neural network model to train the data using enhanced regression algorithm for improving the better and higher accuracy rate by reducing the lower error values. An error rates are calculated using mean square error (MSE) and root mean square error (RMSE) with deep learning techniques.

KeyWords: Soil texture; chemical properties; convolution neural network; linear regression algorithm; mean square error; deep learning.

DOI Number: 10.14704/nq.2022.20.7.NQ22033 **NeuroQuantology2022; 20(7):275-282**

Introduction

Soil texture prediction is one of the major things to maintain the higher productivity of crops and capacity of water stored under soil are highly noticeable matters for farming in different region with different chemical properties present in the soil. Farmers in various regions should know the detailed information about the soil texture to use precision farming strategy for better optimization of profits. The relationships of soil, properties and so are very important for understanding the process of farming. Important features are extracted from the huge amount of remote sensing data which is required for smart analysis techniques. Classification of images for data extraction are very tough process to classify the soil texture using satellite images.

To overcome such critical process, artificial intelligence (AI) and machine learning techniques are worked together for smart analysis of classifying the soil texture using neural networks. The convolution neural network helps to classify the data from huge datasets with artificial neural network (ANN) for pattern recognition which is processed with two different stages as data received as input will be processed with internal activation function and output will be sent to transfer function to determine the neurons will execute the output data or not.

Corresponding author: K. Anandan
Address: [1]Ph.D Research Scholar, PG and Research Department of Computer Science, Chikkanna Government arts college, Tiruppur, Tamilnadu, India, [2,3]Assistant Professor, PG and Research Department of Computer Science, Chikkanna Government arts college, Tiruppur, Tamilnadu, India
E-mail: anandmca07@gmail.com

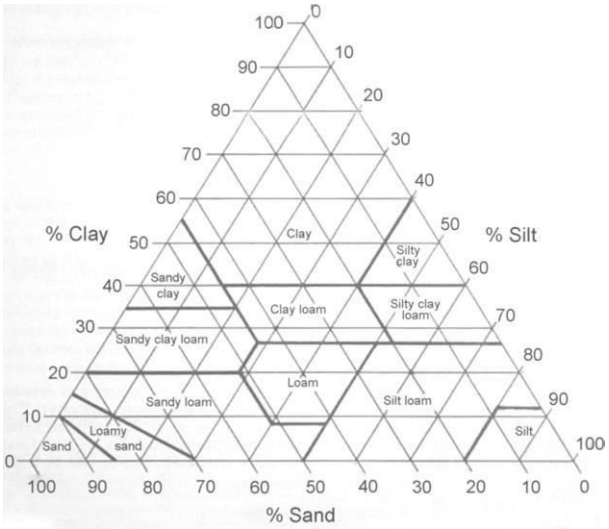


Figure 1. Soil texture categorization triangle chart

There are many neurons are connected with each other in layers to receive input in the layer and the data will be process through more than one hidden layer and produce an output as pattern in a layer.

An enhanced linear regression algorithm is proposed to increase the accuracy and error rates are reduced using deep learning methods.

Artificial neural network is the highly processed technique with robustness, faster processing, general capability, it can easily process with high data spaces. It has capability for non- linear functions to classify the data with non- parametric regression for decision making. It is crucial part using multispectral image which does not have any knowledge regarding distribution of the data, it has supervised learning method such as feed forward back propagation networks to remove the noisy data in the image taken as input data. Pre-processing is the important process to classify the data for data extraction with spatial and spectral information using artificial neural network from LANDSAT images. Also, post- processing the executed results from training neural network is more important to obtain the higher accuracy and performance while using with image processing and classification techniques are the issues. It is difficult to train the model with classifier to learn the classification rule by applying the classifier models in effective methods. It helps to allow the users to use simple classifiers in divide and conquer methods than complicated models. Such methods help to execute the minimum error rate with higher accuracy.

Objective

An objective of this research was to predict the soil texture using selective chemical properties for better execution of accuracy using artificial neural network- based methods by implementing the LUCAS datasets. Chemical properties selected for implementation such as organic carbon, cation exchange capacity, nitrogen, pH level of water, clay, sand, potassium, and phosphorous for effective execution.

Related Works

In this section, the research papers published related to soil texture prediction are reviewed briefly and thereafter we focus on convolution neural network to understand about the merits of the different approach.

Prediction of soil texture is one of the major factors in environmental process. There three different models are randomly used for the prediction of soil texture as CNN, Lucas CNN and ResNet for identification of soil type [1]. For in depth pre-processing of the layers Lucas Coord conv method was used using CNN approach by classifier method and achieved better performance with normal accuracy with some error rates. Prediction of texture helps to classify the crop suit for certain region of area based on some chemical factors and water flow. The datasets are collected from the real time image captured using cameras and it processed for pre- processing methods to extract the information about objects present in the images. Feature vector is the process of categorizing the soil types using some elements like Hue, saturation, value (HSV) with various methods like histogram, wavelet processes, classification

models using linear kernel [2], [3], [4].

Deep learning is one of the methods which helps to solve the problems faster using convolutional neural network (CNN) to predict the soil features [5] using spectral data about the soil is processed with 2D spectrum to present the functions of wavelength and frequency of the data. Network architecture of CNN helps to predict the different soil features by training the networks with CNN model. The max pooling layers are learned about the structure of soil by different properties. The enhanced CNN model can learn the features of pre-processing to neglect the extra processing methods. The image captured using smartphone is implemented with CNN model to predict the texture and structure of the soil in paper [6][7], the low-level image features like color and texture are extracted using existing data about the land.

Using machine learning techniques and algorithms the prediction of soil types is identified based on the hydrological group and it will be trained for classification with four different factors of soil such as sand, slit, clay and conductivity of hydraulic in [8][9].

Based on the parameters of chemical factors in soil helps to predict the soil texture to identify the better profitable crop for farming. These methods made farmers or agriculturists to decide about the particular crop for farming based on the season [10]. All the farmers should have the awareness about which farming does not suit at which season to avoid the losses and for better yielding of cultivation. The future agriculture will be connected with smart devices for easy understanding about the farming and soil properties, etc. using various integrated sensors.

Agriculture is the most important thing and it is considered as pillar of our nation and major of the Indian economy is based on the farming. IoT enabled smart devices with integrated farming system will be helpful more with sensors for better prediction and control the major process over internet which is consist of different hidden layers for enhanced outcomes [11]. Also, the method called multi class classification used in farming for monitoring the farming suitable.

Plantation of crops is important process in agriculture, but before planning for farming it is necessary to prepare the soil for better cultivation and yielding. The soil prediction for cropping was identified and prepared by the farmers with their own interest and knowledge. But now a days, the soil texture and soil variables are calculated by prediction methods using lot of prediction and classification techniques using color image processing. The pH level of different soil images is captured for pre- processing methods and pH factor of each picture with noises are determined [12]. Using different classifiers and prediction methods applied with the input image for prediction of texture and RMSE [13].

Examining the soil properties helps to get better execution of harvests and different AI methods are executed for smart prediction of information about the farming and soil texture prediction [14]. There are major artificial intelligence strategies are followed for smart processing of the prediction such as K-Nearest Neighbour (KNN), Multinomial Logistic Regression, Naïve Bayes Classifier, Artificial Neural Network (ANN) and Random Forest models applied in the data taken as input for processing [15]. Additionally, Naïve bayes theorem, Adaboost algorithm, SVM calculations are integrated methods to improve the precision of output execution.

Soil texture model

proposed the diffuse reflectance spectroscope model to predict the texture using VIS-NIR with MIR and integrated spectra data. Their research executes the correlation of limited band range for different soil properties. The used PLSR method for analysing the soil properties with different data by also implementing PLSR method to predict the soil properties such as sand, silt and clay content with MIR, VIS model. The PLSR model helps to improve the covariance between the data and the response variable. The machine learning algorithm called cubist method are used to build the tree model with integrated classification and linear regression tree model for better prediction of the model.

Statistical methods such as Multivariate Relationships (MARS) have also been seen in the literature for predicting soil texture. It is the process in data mining for the data with no knowledge and the author stated that they taken sample data greater than 5000 and accuracy with cubist and PLSR models not increased.

Convolution neural network model

CNN is one of the parallel processing methods which is a branch of deep neural network learning model with more layered using neural network models. It was introduced in the year 1980s and it requires huge sized data for training with high level

resources and it learn the features of information through back propagation model using multiple layers such as convolution, pooling and connected layers. All layers have their own filters which can extract the features of data without any pre-processing techniques.

Pooling is the operation done on maximum value at each set of feature map which executes the down sampling and it will highlight the important features and pooling layers with average value also applied with feature map and placed between the two layers of convolution with previous layers to apply the pooling operation on each layer of feature maps.

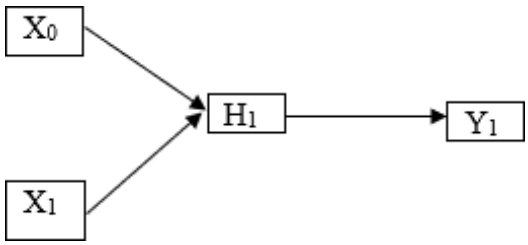


Figure 2: Simple neuron network

The simple component of complete network of neurons are derived with a sum of weights from input data by applying the linear function and activation function is applied to weighted sum are explained in fig 1.

Fully connected network

the fully connected layers of network have neurons for every layer and it receives the input from all the neurons of existing layers which is connected deeply and forms a matrix vector using parameters.

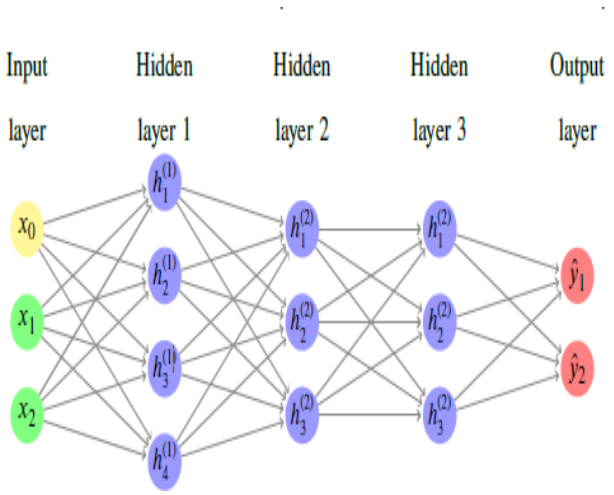


Figure 3. Fully connected convolutional layer

If the hidden layer 3 is considered as convolutional layer, then the input data of previous layer is convolved by the filters to execute the output map of layer 3 [27][28]. If the pooling and sub sampling layer l is given as m1 = 4, where m is the feature map of previous layer. The output feature map of each unit is represented as average within the boundary of corresponding layer (l – 1) as feature map.

Existing Proposed Research

This section explained about the previous research implemented for predicting the soil texture properties using CNN network approach and hybrid optimization algorithm using selective chemical properties as parameters.

Phase 1

The phase one research was implemented with convolutional neural network model to predict the soil properties for hyperspectral data. CNN is trained for better understanding the pattern of processing the hyperspectral data. The 6 soil properties are predicted in this phase such as organic carbon content, cation exchange capacity, nitrogen content, pH level, clay and sand particles [31]. The soil texture prediction is based on the presence of silt, sand and clay particles, the CNN is trained with hyperspectral data as input with multiple arrays and the model performed to evaluate the RMSE. Deep learning approach was proposed by representing the non-linearity for scalable data. Finally, concluded with average error rate for all the six different types of soil properties as 5.68%.

Phase 2

The phase two was implemented with convolutional neural network model integrated wit hybrid optimization algorithm called Levenberg Marquardt backpropagation optimization algorithm. The proposed work was enhanced with two more soil properties were added as eight total soil properties such as potassium and phosphorous to provide a novelty of research and predict the texture of soil by reducing the error rate as 2.6% compared with phase one error value and increase the accuracy of prediction as 96.8% [32].

Proposed Methodology

To evaluate the performance of convolution

network model to predict the soil texture the chemical properties as parameters are used selectively and the statistics model was shown in the table 1 which executes the better results.

Table 1. Training statistics using CNN using 100 bootstrap realisations

	Test data		Training data	
	Mean error	RMSE	Mean error	RMSE
N	2.17	9.43	2.4	1.92
P	0.67	3.40	0.33	2.50
K	7.89	4.89	6.12	3.56
pH	0.98	0.72	0.67	0.54
CEC	1.85	5.12	1.24	4.35
OC	1.49	41.2	1.2	34.21
Clay	-0.96	12.76	-0.57	9.28
Soil	2.73	7.34	1.86	6.51

Linear regression is the basic machine learning algorithm used to find the linear relationship between the target and one or more predictors. Linear regression has two types such as,
Simple linear regression
Multiple linear regression.
The proposed research is implemented with simple linear regression algorithm for classifying the best fit regression relationships of data points.

Best fit lines

The regression line is called as line of best fit due to its line drawn through the data points are more accurate and the line can minimize the distance of actual scores from the distance of predicted score values. The below equation is used for modelling the best fit line using the linear regression.

$Y = a\_0 + a\_1 * X$ (1)

The best fit line of linear regression is also described as,

$Y = bX + a$... (2)

Where, the 'b' indicates the slope of the line and 'a' indicates the intercept, which means the value of Y, when the value of X = 0.

Cost functions

Cost function helps to figure the best possible values for a_0 and a_1 which will provide the best fit line for the data point. To minimize the worst fit lines that is error line the below given equation used to avoid worst fit lines,

Minimize, $\mathbf{1}\sum_{\boldsymbol{n}}(\boldsymbol{predi}-\boldsymbol{yi})\mathbf{2}$.....(3)

$$J = \mathbf{1}\sum_{\boldsymbol{n}}^{\boldsymbol{i}=\mathbf{1}}(\boldsymbol{predi}-\boldsymbol{Yi})\mathbf{2} \quad (4)$$

Gradient descent

This is the best method to update the a_0 and a_1 to reduce the cost function that is mean square error (MSE).

Linear regression:

$$A0 = a0 - \alpha.\ \mathbf{2}\sum_{\mathbf{n}}^{\mathbf{i}=\mathbf{1}}(\mathbf{predi}-\mathbf{Yi}) \quad (5)$$

$$A1 = a1 - \alpha.\ \mathbf{2}\sum_{\mathbf{n}}^{\mathbf{i}=\mathbf{1}}(\mathbf{predi}-\mathbf{Yi}).\mathbf{i} \quad (6)$$

Gradient descent used in deep learning that is machine learning model is to find the best parameters to reduce the cost function of the model. It helps to solve the optimization issues with prior iterations and it find the local minimum functions and widely used in machine learning methods.

Results And Discussion

Soil texture prediction can be processed with data collected from the database called LUCAS with 12,000 attributes and the entire datasets are taken after cleaning process for training the neural network and selectively eight chemical and soil properties are considered for texture prediction. The eight properties are organic carbon, pH level of water, potassium, phosphorous, carbonate content, nitrogen value, silt and clay particles. The hyperspectral data are considered for training the network for better prediction process.

Convolution neural network is built to train the huge dataset for effective accuracy improvement and error rates are calculated by the RMSE and ME using the below given equation,

The below figure explained about the overall process of the proposed research and algorithm process,

$$\text{MeanError} = \frac{1}{\boldsymbol{n}}(\boldsymbol{Yi}-\boldsymbol{Yi})\mathbf{2} \quad .....(7)$$

$$\text{RMSE} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i}=\mathbf{0}}^{\boldsymbol{ni}=\mathbf{0}}(\boldsymbol{Yi}-\boldsymbol{Yi})\mathbf{2}} \quad .....(8)$$

Linear regression method is the best analysing process to identify the variables which can perform the determination of certain factors and it helps to ignore the factors which are unnecessary for training the data while testing.

It is one of the machine learning algorithm which is the dependent variables and supervised learning method helps to predict the factors as dependent

target based on the given datasets variable.

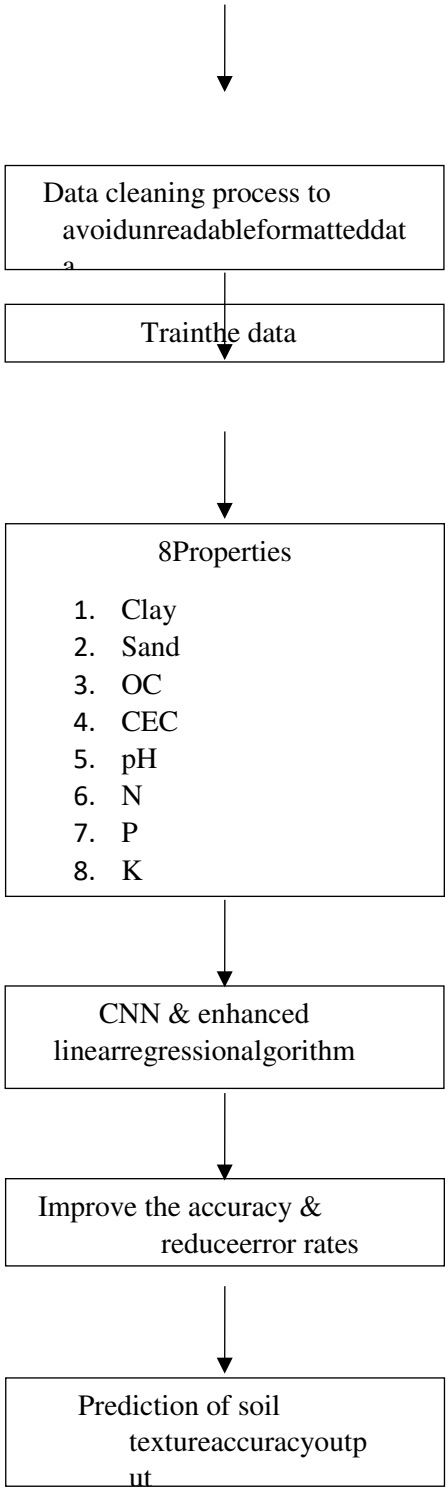


Figure 4. Overall workflow of proposed model

Proposed linear regression algorithm

Step 1:	Feed the input
Step 2:	Read the entire data (n) given for processing.
Step 3:	For i = 1 to n, read Xi and Yi then i.
Step 4:	Initialization of training process
Step 5:	Sum values are calculated, for i=1 to n, next i.
Step 6:	Constant value of a and b of y = a + bx is calculated. b = (n * sum XY – sum X * sum Y)/ (n*sumX2 – sum X * sum X) and a = (sum Y - b*sum X)/n.
Step 7:	Value of a and b executed
Step 8:	Stop the process

The simple linear regression algorithm mentioned by the equation,

$$Y = a + bx \quad (9)$$

Enhanced linear regression algorithm

An enhanced linear regression algorithm using least square method helps to predict the best fit of the datapoints and avoid the worst fit faster. The proposed algorithm works much better than the normal linear regression algorithm by minimizing the sum of square errors between the data and the model implemented

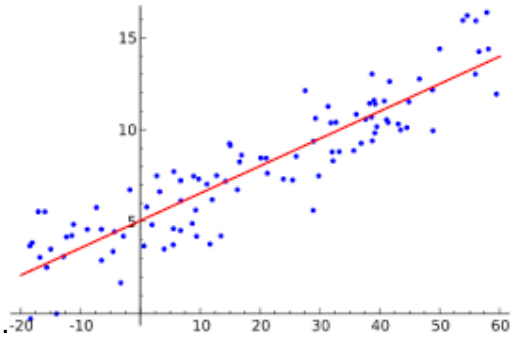


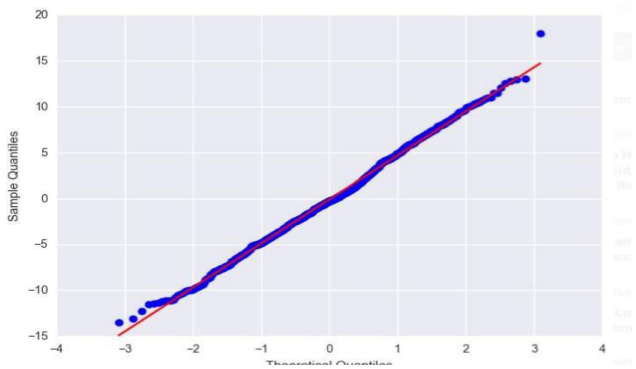
Figure 5. Best fit- Linear regression model

The enhanced linear regression algorithm is applied on simple LR algorithm using least square method in three steps as normal equation formed and solve it simultaneously for the values a and b

and substitute the values of a and b with equation 9 for best fit line.

Normal Distribution applied

Normal distribution (ND) is the process of displaying the data near to the mean frequently than the data far away from the mean value. The normal distribution also known as gaussian



distribution.

Figure 6. Normal distribution of LR

Gaussian distribution is known as skew and it will be calculated using python with Scipy module. The below figure shows the normal distribution which can helps to increase the accuracy of linear models. Here, the box- cox transformation method makes the skew to be fixed upto 80% of cases and it is very useful and powerful to tool to fix the value of skewness which makes the non- normal distribution fit to normal distribution fit.

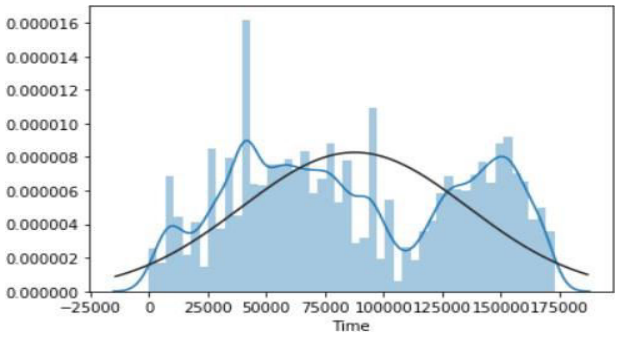


Figure 7. Gaussian distribution value before transformation

Table 2. Comparison of SLR & enhanced LR

S. No	Simple linear regression	Enhanced linear regression
1	It does not have any improved techniques	Box- cox transformation applied
2	It has a chance to execute low accuracy	It helps to increase the lower accuracy to higher
3	It does not avoid any unfit data points	It will avoid the far away data points and make the best fit for error reduction.

Table 3. Comparison results of existing and proposed algorithms.

	Algorithm	Accuracy	Error rate
Existing methods	CNN	95	5.68
	VNIR_SWIR	95.2	7.7
	SOM	87	6
	FFNN	92	3-6
	LMO	96.7	2.6
Proposed method	CNN + enhanced linear regression model	97.4	2.1

Conclusion

Prediction of soil texture using their chemical properties is really a tough process using images, to overcome the situation CNN model was developed to train the data collected from the LUCAS database as hyperspectral data. There are eight different soil properties are taken as parameters to predict the texture of soilsuch as organic carbon, cation exchange capacity, nitrogen, pH level, clay, sand particles, potassium and phosphorous. The proposed research work is implemented using convolutional neural network (CNN) with enhanced linear regression algorithm for better prediction of

texture with higher accuracy as 97.4% and lower error rates.

References

Humaidi, et.al., "Review about deep learning concepts, applications and challenges", journal of big data, pp. 1-7, 2021.

Wetterlind and Sternberg, "Enhanced method to predict the performance by VIS-NIR spectroscopic soil using data mining", journal of European soil science, pg. 718-729, 2017.

P.Bharati, et.al., "Review about the Deep learning techniques for RCNN masking the data", journal of computational intelligence in pattern recognition, springer, pp. 657-668, 2020.

coblinski, et.al., "Soil texture prediction through various wavelength regions at different soil depths", journal of soil science, pg 189-193, 2020.

L. Deiss and Chang, "SVM tuning regression model to improve the accuracy of predicting the soil properties", Geoderma 365, 2020.

Devlin, et.al., "Pre-training of deep bi-directional of transformers for the understanding of different languages", arXiv: 1810.04805, 2018.

Geng, Z.,Wang, Y, "Automated design of CNN with multi-scale filters for cost-efficient seismic data classification", Journal of Nature communications, pp. 1–11, 2020.

Hertel, L., Collado, J., Sadowski, "Robust hyperparameter optimization for machine learning", SoftwareX 12, 100591, 2020.

Hira, Z.M., Gillies, D.F., "A review of feature selection and feature extraction methods applied on microarray data", Journal of Advances in bioinformatics 2015.

Hira, Z.M., Gillies, D.F. "A review of feature selection and feature extraction methods applied on microarray data", Journal of Advances in bioinformatics, 2015.

Hu, W., Huang, Y., Wei, L., Zhang, F., Li, H. "Deep convolutional neural networks for hyperspectral image classification", Journal of Sensors, 2015.

Ioffe, S., Szegedy, C. "Batch normalization: Accelerating deep network training by reducing internal covariate shift", International conference on machine learning, PMLR. pp. 448–456, 2015.

Jaderberg, M., et.al., "Population based training of neural networks", arXiv preprint arXiv:1711.09846, 2017.

LeCun, Y., Bengio, Y., Hinton, G. "Deep learning", Nature, vol. 521, pg. 436– 444, 2015.

Lee, J.G., Jun, S., et.al., "Deep learning in medical imaging: general overview", Korean journal of radiology, vol. 18, pg. 570–584, 2017.

Liu, L., Ji, M., Buchroithner, M., "Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery", Journal of Sensors, vol. 18, 2018.

Morellos, A., et.al., "Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using vis-nir spectroscopy", Journal of Biosystems Engineering, vol. 152, pg. 104–116, 2016.

Ng, W., Minasny, B., Malone, B., Filippi, P., "In search of an optimum sampling algorithm for prediction of soil properties from infrared spectra", Journal of data science, 2018.

Ng, W., Minasny, B., Mendes, W.d.S., Demattê, J.A.M. "The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data", Journal of Soil science vol. 6, pg. 565–578, 2020.

Otter, D.W., Medina, "A survey of the usages of deep learning for natural language processing", IEEE Transactions on Neural Networks and Learning Systems, vol. 32, pg. 604–624, 2020.

Padarian, J., Minasny, B., McBratney, A. "Using deep learning to predict soil properties from regional spectral data" Journal of Geoderma Regional, 2019.

Riese, F.M., Keller, S. "Soil texture classification with 1d convolutional neural networks based on hyperspectral data", arXiv preprint arXiv:1901.04846, 2019.

Rossel, R.V., et al., "A global spectral library to characterize the world's soil", Journal of Earth- Science Reviews, vol. 155, pg. 198–230, 2016.

Ruder, S. "Multi-task learning objectives for natural language processing", URL: http://ruder. io/ multi- task- learning-nlp/(visited on 10/02/2018), 2017.

Ruder, S. "An overview of multi-task learning in deep neural networks" arXiv preprint arXiv:1706.05098, 2017.

Santana, et.al., "Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters",

International journal of Spectro Molecular and Biomolecular Spectroscopy, vol. 191, pg. 454–
462, 2019.

Shen, Z., Rossel, R.V. "Automated spectroscopic modelling with optimised convolutional neural networks", Journal of scientific reports, vol. 11, pg. 1–12, 2021.

Sun, Y., et.al., "Completely automated CNN architecture design based on blocks", IEEE transactions on neural networks and learning systems 31, 1242–1254, 2019.

Yang, M., Xu, D., Chen, "Evaluation of machine learning approaches to predict soil organic matter and ph using vis-nir spectra", Journal of Sensors, vol. 19, pg. 263-270, 2019.

Xu, H., Xu, D., et.al., "Rapid determination of soil class based on visible-near infrared, mid- infrared spectroscopy and data fusion", Journal of remote sensing, vol. 12, 2020.

K.Anandan, et.al., "Convolutional neural network approach for predicting the soil texture properties", Indian journal of science and technology, vol. 14, issue. 3, pg. 190- 196, 2021.

K.Anandan, et.al., "The Prediction of Soil Texture Properties using Hybrid CNN-LMO Algorithm", International journal of mechanical engineering, vol. 7, issue 4, Apr. 2022.