



Supervised Ontology Oriented Deep Neural Network to Predict Soil Health

Kushala Vijaya Kumar Mummigatti^{1*}, Supriya Maganahalli Chandramouli²

¹ Sri Siddhartha Academy of Higher Education, Tumkur 572107, Karnataka, India

² Department of Information Science, Sri Siddhartha Institute of Technology, Tumkur 572105, Karnataka, India

Corresponding Author Email: vmkushala@gmail.com

<https://doi.org/10.18280/ria.360220>

ABSTRACT

Received: 6 March 2022

Accepted: 15 April 2022

Keywords:

ontology, deep learning, neural network, backpropagation, scale conjugate gradient, soil health, classification

Soil health plays a vital role in agriculture. A nutrient-rich soil helps in better crop growth and high yield. The agriculture data in India is haphazard and no major effort is seen in maintaining them. Soil chemical property is a basic knowledge to decide on cultivation. Knowledge base to help farmers analyse the soil health by using the chemical properties as the main feature in predicting the health and quality of the soil before the cultivation is a key factor for a better production result. This study drives the idea of building a domain ontology model for soil and also utilizes a neural network in predicting the soil by classifying it as healthy or unhealthy based on six chemical parameters that explain the property of soil. Ontology plays as a knowledge base in storing the properties of the soil which also helps in enabling artificial intelligence concepts on the knowledge to make better decisions. MATLAB deep learning toolbox is used to implement the classification and also TensorFlow's Keras was used to handle the data pre-processing, normalization and also the network architecture to validate the result from the toolbox. MATLAB employs the Scaled Conjugate Gradient algorithm and performs with 92% accuracy in achieving the classification of soil.

1. INTRODUCTION

Globalization made a tremendous impact on India which impacted many sectors to grow into a better state. But agriculture being age-old employment in India was not much polished with technologies and stood as a neglected field with poor performance delivery. With a highly improved, timely and nutrient production of crops, the country can establish better trade culture, provide employment opportunities in the agriculture sector, and increase the technologies involved in the process, which directly helps the increase the income and growth of the country. In a developing country with a majority of the population dependent on agriculture as a main source of income, driving agriculture to meet the globalization speed plays a very important role.

Human population dependency to produce high agricultural yield with improper knowledge of chemical fertilizers has an adverse effect on cultivation land and has diminished the fertility of the soil and also the nutrient value of yield. Adoption of new scientific techniques to test the quality of soil in a short time helps the farmer to make a better decision about the fertilizers and the crop to be cultivated in their soil.

India with its main revenue from agriculture has failed in adopting the technology and distributing it to the farmers in a geographical aspect. Agriculture field is the most dynamic, with huge and diversified data. Artificial intelligence has a wide practice in the industry with evidence of delivering good results with diversified data. An attempt is made here in adapting it to the agricultural land. To facilitate the approach, we are adopting an ontology framework which helps in an efficient storage with respected relations among the data helping the data.

The main objective of this paper relies on classifying the soil whether it is healthy soil for cultivation or unhealthy soil. The chemical nutrients it lacks plays an important role to classify them into cultivation and non-cultivation soil. To classify the soil according to our objective: soil has two properties physical and chemical. Physical properties as it indicates involve the "look and feels" of soil i.e., color, texture, structure, porosity, density, consistency, aggregate stability, and temperature measured from external devices. Chemical properties are an important factor in determining the quality and nutrient factors of the soil and it is explained by the test conducted for pH, EC, Potassium, Nitrogen and Phosphorous. These are the major parameters we have considered to validate soil health.

There is a huge dimension to agricultural knowledge. A knowledge base or a framework to preserve the knowledge related to agriculture by invoking many dimensions like soil with chemical and physical properties, crop type, fertilizer type which involves chemical, organic, bio-chemical etc., weather, cultivation type, harvesting time depending on the crop etc., which holds many of these kinds of dimension is lacking, therefore, we come up with ontology that can handle these dimensions effectively. Ontology helps in capturing the knowledge and provides an understanding of the domain. Ontology helps in describing and conceptualizing the basic concepts in the domain used and establishes a relationship among them. This helps in the retrieval of the knowledge from the base effectively with a minimum amount of querying [1].

Indian cultivation land varies widely with different soil types, cultivation practices, weather and crops. Currently, there are many ontologies developed for the domain of agriculture. Name a few of them: AGROVAC, Crop

Cultivation Standards, Crop ontology etc. These ontologies are developed by specific organizations under specific regions, knowledge, interfaces, search criteria etc., explaining, that AGROVAC ontology is widely used but it involves the knowledge base of horticulture and animal husbandry which makes adapting to local regions a gap leading it as a drawback. Crop cultivation standards involve only knowledge of standardizing the crop with no knowledge of considering soil as a parameter. The CropOnto incorporated all the attributes of AGROVAC along with farming activities. Our review of domain ontology extensively proves the reuse of the ontology in Indian farmlands is challenging and fails. They fail in considering the property of soil and a framework to decide on them. Hence, we attempt to build a domain ontology for the soil's chemical properties.

With the domain knowledge present with us which we obtained by constructing ontology, we can apply deep learning techniques to classify the soil into a healthy or unhealthy class which helps in the fast-decision-making process during cultivation decision. Deep learning is a part of artificial intelligence that uses neural network architecture to process information. They are well known for their speed and accurate results as they are conceptualized based on the human brain's working process. Deep learning architecture beats many technical challenges faced by machine learning. It can handle a large amount of data without feature engineering, it efficiently delivers the results compared to machine learning.

We performed the same study by implementing different machine learning algorithms with the knowledge base i.e., by constructing a domain ontology of soil. The maximum accuracy of the model achieved approximately was 80% [2]. Machine learning failed to perform well on the varying data samples collected from us. An efficient algorithms like decision tree, random forest, XGBoost was implemented in classifying the objective. Which resulted in 77,85,80 percent of accuracy along with very poor ROC and precession values. Deep learning can manage any kind of data i.e., data without common patterns as needed by machine learning. With the advantages of deep learning studied, it is evident that neural networks can produce better results in many challenging problems related to agriculture [3]. By adopting deep learning we achieved in a better accuracy of 92%, the problems of machine learning faced in handling dissimilar was easily and timely achieved.

2. LITERATURE SURVEY

Artificial intelligence is playing an evident role in many fields and effectively delivering very good results. A neural network can deliver better results as they are trained with specific data which are processed logically in producing the result [3]. ANN is composed of three layers named as the input layer, the hidden layer and the output layer [4]. The input layer is the only visible layer in the neural network (NN) it acts as a receiver and transformer by accepting the information in various forms and pass on to the next layer. Hidden layers are the magnificent layers that handle the whole computation. It handles many computations works like weight adjustment, adjusting the bias, activation function, transfer function etc., the output layer receives the computational works done by the hidden layer and outputs the prediction as per the user requirement.

The learning phase of the neural network involves adjusting

the weights in the network such that it can predict the appropriate class label. A NN is classified into two types supervised and unsupervised [5]. Supervised learning of NN is trained based on input-output data which are fed to the network, they generally involve multi-layer feed-forward which is trained using a back-propagation algorithm. Unsupervised learning of NN does not rely on input-output data but trains automatically on the information present in input-output example data.

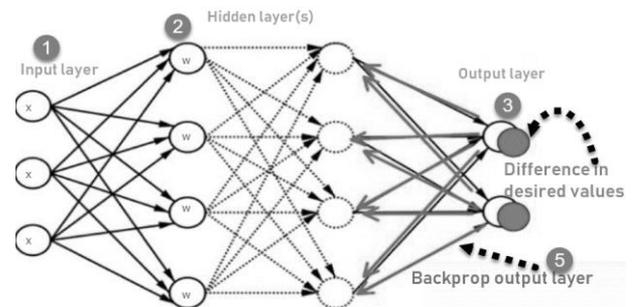


Figure 1. Basic structure of neural network with back propagation

Convolution neural network along with android based was used in soil classification, which consisted of 10000 images of soil from Cagayan Valley, Philippines. The images contained the upper layer images of soil that belonged to different classes. Using CNN with a few layers of hidden layers and learning rate the soil images were able to get classified into 16 different classes with 90% and more accuracy [4].

Soil class prediction using orbital sensing data and terrain attributes derived from digital sources and geology departments using ANN was successful in producing an accuracy of 90 to 95%. A neural network simulator along with a backpropagation algorithm is used for this study [3]. Soil class classification along with soil moisture and temperature profile prediction model was built using neural network backpropagation and Levenberg-Marquardt algorithm on the data collected from remotely sensed microwave [5]. Network architecture was designed with three hidden layers with the sigmoid and tangent transfer as activation functions to classify the soil into its classes.

Collecting the factors like organic matter, essential plant nutrients, and micronutrients required for the growth of a crop was evidently found using the backpropagation algorithm which suggests and finds the correlation between the nutrient facts, and external factors' effects and suggests the growth of the crop [6]. The backpropagation idea was also implemented to classify the soil with the ability to classify 5 different types of soil by considering 11 input parameters [7, 8]. The same idea was used to classify the soil based on its general behavior and the given physical condition, the input parameter is porosity and water content and the output is a type of the soil class it belongs to with 95 to 99 percent accuracy, the similar idea is implemented with different algorithms of deep learning was achieved to classify the soil with similar accuracy level [9-11].

An ontology is a set of concepts and categories in a subject area or domain that possesses the properties and relations between them. They help in providing background knowledge that can be used by machine learning and deep learning models [12]. They also help in training the better model by improving the data quality. Navigation between one structure of data to

another structure of data is easily made with their usage of them. Ontologies are also used as structured output in the task of predicting whether the entity has relationships with one or more classes. Ontology helps as a tool between humans and computers by improving the communication between them by reusing the knowledge, information and data.

Agriculture Online Service (AOS), is a domain ontology developed to provide agricultural knowledge management with other semantic applications. This ontology supported a multi-linguistic approach to storing the knowledge. As the knowledge base for agriculture is a lack in a county like Sri Lanka, a domain ontology was developed with major concern to users. With this approach, they achieved creating a repository of knowledge and retrieval of information according to the query concerning the context requested by the user [13]. Research for domain Ontology for agriculture was conducted in Pakistan which makes the stakeholders store, manage and share the instances effectively and reuse them effectively. Their main intention was to create E-modeling Ontology was agriculture [14].

Farmer helping system ontology developed by Chennai, Indian Researchers concentrated on providing a helping hand framework for farmers [15]. This integrates different information from soil, pesticides, crop, methodology and many more. The framework accepts the user requests and analyzes it to reach the information to the appropriate knowledge base with the solution. The data is semantically annotated which helps in storing, analyzing and retrieving the data.

With these studies considered and reviewed artificial intelligence and ontology are playing an important role in different aspects of agriculture. Deep learning algorithms have been extensively used in making decisions on different prospects. The importance of knowledge in aspect towards a particular domain of agriculture can be addressed with ontology.

3. METHODOLOGY

As discussed previously with the architecture of NN concerning Figure 1, the backpropagation technique has three layers, input hidden and output layers. Backward propagation of errors in the technique adopted by the backpropagation algorithm. This is, after every forward pass in the network this algorithm does a backward pass to adjust the weights and bias in the neural network until the errors are minimized. The first layer takes the input parameters which can be either simple scalar or complex vector forms of data. This advantage is completely leveraged in building our soil classification on the devise data.

The data was collected from soil testing institutes around Mysore district, Karnataka State, India. The data was handwritten in the logbooks, these are then typed into the computer. The data contained the parameters like soil type, acres, survey number of the land, owner information, PH, EC, Nitrogen, Phosphorus, and Nitrogen values. We discarded the values like owner information, acres, and survey number.

Observations and handling of the data: Basic statistics of the data project that 80% of the soil data appeared to be unhealthy and 20% is a healthy class. This alone proves the data we had gathered dominated class-imbalance characteristics. The data did not project any similarity or inferences. The value of each parameter varied drastically from each acre which we inferred

from the pair plot. By consulting the soil experts, we derived the desired values of each parameter that decide the health of the soil. With the derived value, a new feature called as the target is added representing 0 as unhealthy and 1 as healthy soil.

As we had already mentioned the challenge in agricultural data is there are no similar patterns in the data. The data varied drastically with very few changes in geographic region and cultivation practices. The soil properties changed in every single acre of land due to different cultivation practices, fertilizers use and the crops produced. Class imbalance with the dominance of unhealthy soil is one of the challenges which was not effectively handled by the machine learning approach.

ph	ec	nit	phos	potas	target
6.41	0.019	211.0	9.51	101.76	0
7.70	0.210	348.0	21.70	74.75	0
6.74	0.103	285.0	14.70	143.51	0
7.01	0.071	234.0	27.41	137.45	1
6.75	0.091	274.0	12.75	117.41	0

Figure 2. The raw data collected from soil test institutes (Visualization from Jupyter Notebook)

Figure 2 is the raw data collected from the agricultural soil test center. Agricultural data is not digitalized. The data was loaded in the Jupyter notebook to understand the relationships using EDA and to pre-process.

Figure 3 represents the architecture and workflow adopted in classifying the soil using domain ontology. The data is collected from the soil testing centers and also acquired in-depth knowledge of the soil domain. This helps in identifying the main features needed for the knowledge base and how to obtain the relation between them. Construct a soil domain-oriented ontology. Validate the ontology with new data. With the help of ontology, the framework applies deep learning neural network to classify the soil to the desired class.

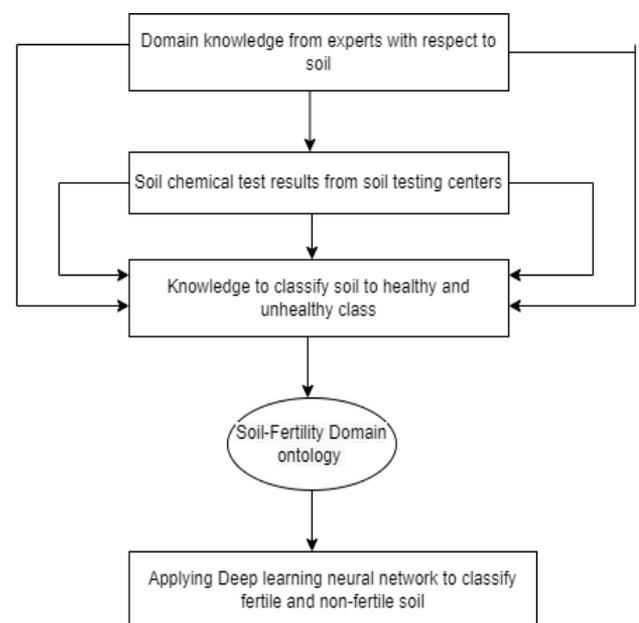


Figure 3. Methodology adopted frame work

3.1 Domain ontology

Ontology is developed using Protégé 4.3. This is one of the major steps involved in understanding the significant class features and their inferences. This creates domain-specific knowledge representing the soil class and its properties. The important reason to generate ontology is to acquire and get a profound insight into the features in the dataset collected and treated. The domain ontology with the combination of chemical property and soil health as the class and their respected attributes are visualized using asserted class hierarchy visualization tool in Figure 4. The hierarchical structure between the class and their relation is depicted as tree structure in Figure 5.

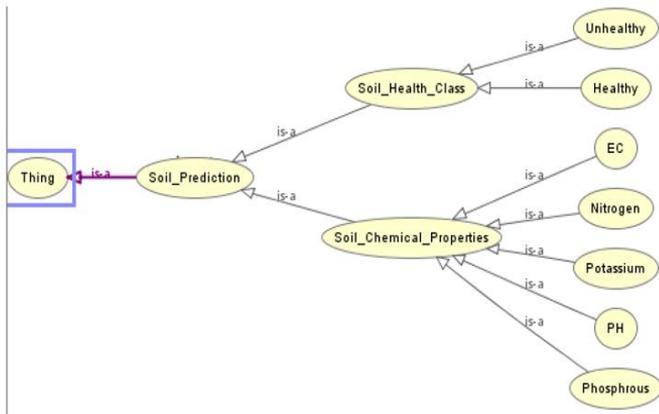


Figure 4. Asserted class hierarchy of soil health

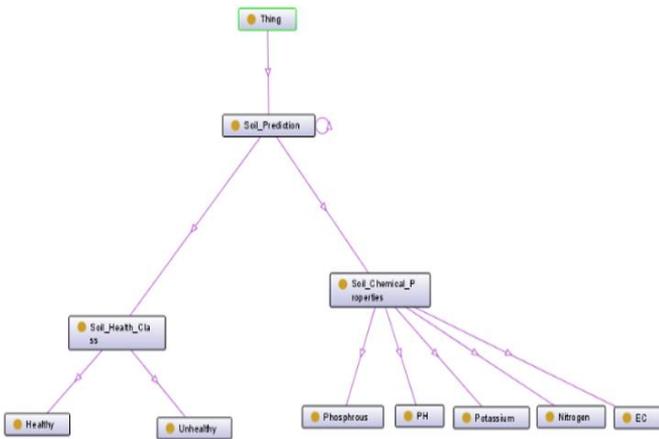


Figure 5. Tree structure produced by protégé

Using MATLAB, the collected data was trained using scaled conjugate gradient algorithm. The same data was also trained using TensorFlow’s Keras to re-validate the accuracy and performance of the model. The network architecture was composed of five hidden layers which employed batch-normalization, dropout layer with learning rate of 0.05 to improve the performance stochastic gradient descent was also adopted.

3.2 Scaled conjugate gradient algorithm (SCG)

SCG algorithm is proved to outperform the bench mark of back propagation algorithm, it is completely an automated and less time-consuming algorithm as it avoids linear search process to adjust the weights [16]. In SCG the search is

performed along with conjugate directions which makes faster convergence other than the steepest descendent algorithms.

Eq. (1) represents the generalization of forwarding pass. Error is calculated using Eq. (2) and performs Backpropagation. In SCG step size function is a quadratic approximation of the error function which makes it more robust and independent from the user.

$$h1 = wn * i1 + wn+1 * in+1 + b1 * 1 \quad (1)$$

$$\text{Error} = \text{sum} * \frac{1}{2} (\text{target} - \text{output})^2 \quad (2)$$

$$\text{SCG} = E (wk + \sigma kp) - E (wk) / \sigma k + \lambda kp \quad (3)$$

4. RESULTS AND DISCUSSION

The data from agricultural test institutions are trained using a neural network that was stored in our domain ontology. The predictions are 0 and 1, 0 being unhealthy and 1 being healthy class. The accuracy of the classification model is measured based on the confusion matrix and ROC curve.

Figure 6 represents the overall accuracy of the model which is 90.9%. The accuracy of the data set which was split into training, test and validation is also present. Figure 7 represents the ROC curve which is used to analyze the accuracy of the classification model. This is used to measure the degree and capacity of the model in achieving the classification of the classes.

The results discussed above evidently prove a better performance and knowledge base compared to the study we made in our survey. The study proved that there is a lack of knowledge base for the easier retrieval of knowledge. With the domain ontology built we defend to have a knowledge base. With easy access to knowledge neural networks performed better in classifying the soil.



Figure 6. Confusion matrix for the soil data-set classification model

Figure 8 represents training vs test loss in every epoch. The figure tells us that error decreased and validated at 0.0631 at 31 iterations which is leading to a better performance.

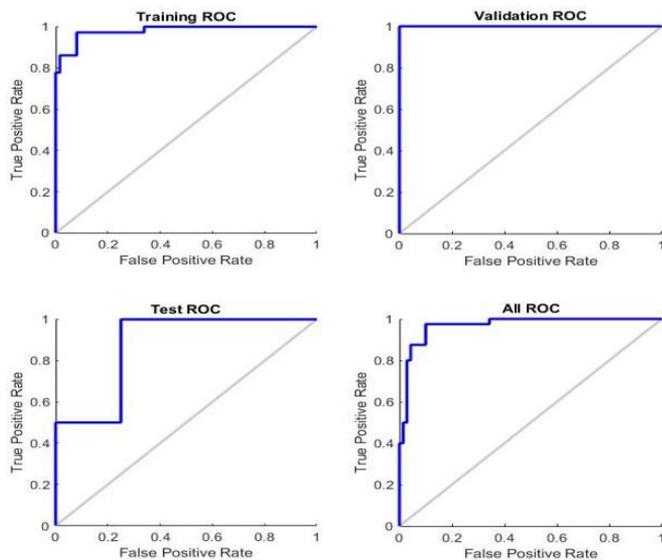


Figure 7. ROC curve for the classification model

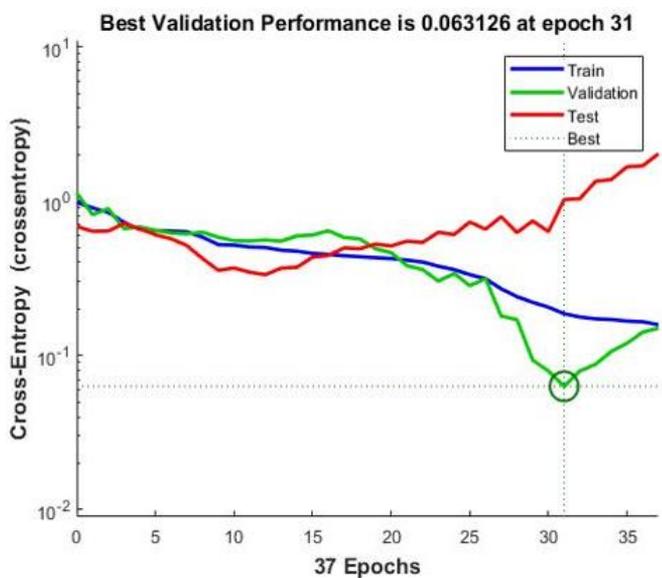


Figure 8. Training vs Test loss

5. CONCLUSION

Agricultural data is rapidly changing data. The data varies drastically from a geographical region, crops grown, fertilizers used, cultivation methodology and various other external factors. Agriculture is the field in India that lacks digitalization of the data, knowledge management system or storage system. The agriculture data has many interrelated dimensions. Storing and managing such kinds of data need to be available according to the geographical region. This type of system will help in taking decisions by minimizing the time by farmers in deciding the soil health.

With this intention, a domain-oriented ontology of soil is built by identifying the concepts and their relations in soil. Which helps in storing the knowledge soil properties as a knowledge base. A decision support method was also attempted by using a deep neural network to classify the soil. Classification of the soil will help the farmers to treat the soil with proper nutrients. We have achieved the knowledge base and also the classification with 92 percent of accuracy. The

ontology can be extended with other factors like climate, fertilizers, and crops and adopt an artificial intelligence method to get the desired result.

Predicting whether the soil is healthy or unhealthy is an initial step in the decision of the cultivation. Along with chemical properties external factors like humidity, and temperature using external devices can also be adopted using wireless sensors, and IoT devices to decide on the soil quality to be more confident. Soil images data can also be an additional consideration to classify the soil.

From the above implementations, we can conclude that Deep learning and Ontology combined can result in evident, easeful results in the agricultural field.

REFERENCES

- [1] Taye, M.M. (2010). Understanding semantic web and ontologies: Theory and applications. arXiv preprint arXiv:1006.4567.
- [2] Kushala, V.M., Supriya, M.C., Suma, N.R. (2021). Supervised machine learning technique to predict soil health. Turkish Online Journal of Qualitative Inquiry (TOJQI), 12(7): 1622-1630.
- [3] Kuwata, K., Shibasaki, R. (2015). Estimating crop yields with deep learning and remotely sensed data. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 858-861. <https://doi.org/10.1109/IGARSS.2015.7325900>
- [4] Atluri, V., Hung, C.C., Coleman, T.L. (1999). An artificial neural network for classifying and predicting soil moisture and temperature using Levenberg-Marquardt algorithm. In Proceedings IEEE Southeastcon'99. Technology on the Brink of 2000 (Cat. No. 99CH36300), pp. 10-13. <https://doi.org/10.1109/SECON.1999.766079>
- [5] Calderano Filho, B., Polivanov, H., Chagas, C.D.S., Carvalho Júnior, W.D., Barroso, E.V., Guerra, A.J.T., Calderano, S.B. (2014). Artificial neural networks applied for soil class prediction in mountainous landscape of the Serra do Mar¹. Revista Brasileira de Ciência do Solo, 38(6): 1681-1693.
- [6] Lagarteja., G.J. (2020). Android-based soil series classifier using convolutional neural network. International Journal of Scientific & Technology Research, 9(2): 2277-2285.
- [7] Odhiambo, L.O., Freeland, R.S., Yoder, R.E., Hines, J.W. (2002). Application of fuzzy-neural network in classification of soils using ground-penetrating radar imagery. In 2002 ASAE Annual Meeting (p. 1). American Society of Agricultural and Biological Engineers.
- [8] Ghosh, S., Koley, S. (2014). Machine learning for soil fertility and plant nutrient management using back propagation neural networks. International Journal on Recent and Innovation Trends in Computing and Communication, 2(2): 292-297.
- [9] Yee, K.M., Aung, T.Z., San, T. (2019). Soil type classification based on neural network. International Journal of Creative and Innovative Research in all Studies (IJCIRAS), 2(3): 6-11.
- [10] Htun, W., Htay, S. (2010). Classification of soil type using backpropagation neural network (Doctoral dissertation, MERAL Portal).

- [11] Elarabi, H., Ali, K. (2009). Soil classification modelling using artificial neural network. The International Conference on Intelligent Systems (Icis2009), Kingdom of Bahrain.
- [12] Xiong, J., Yang, Y., Yang, Z., Wang, S. (2010). An online system for agricultural ontology service. In 2010 Third International Conference on Intelligent Networks and Intelligent Systems, pp. 479-481. <https://doi.org/10.1109/ICINIS.2010.106>
- [13] Walisadeera, A.I., Ginige, A., Wikramanayake, G.N. (2014). User centered ontology for Sri Lankan agriculture domain. In 2014 14th International Conference on Advances in ICT for Emerging Regions (ICTer), pp. 149-155. <https://doi.org/10.1109/ICTER.2014.7083894>
- [14] Ahsan, M., Motla, Y.H., Asim, M. (2014). Knowledge modeling fore-agriculture using ontology. In 2014 International Conference on Open Source Systems & Technologies, 112-122. <https://doi.org/10.1109/ICOSST.2014.7029330>
- [15] Shyamaladevi, K., Mirnalinee, T.T., Trueman, T.E., Kaladevi, R. (2012). Design of ontology based ubiquitous web for agriculture-Aa farmer helping system. In 2012 International Conference on Computing, Communication and Applications, pp. 1-6. <https://doi.org/10.1109/ICCCA.2012.6179214>
- [16] Møller, M.F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4): 525-533. [https://doi.org/10.1016/S0893-6080\(05\)80056-5](https://doi.org/10.1016/S0893-6080(05)80056-5)