Predicting fertilizer treatment of maize using decision tree algorithm

Nusrat Jahan[1], Rezvi Shahariar[2]
[1]Daffodil International University, Dhaka, Banglasdesh
[2]Institute of Information Technology, University of Dhaka-1000, Banglasdesh

Article Info	ABSTRACT
	Machine learning approaches are progressively successful in image based analysis such as different diseases prediction as well as level of risk assessment. In this paper, image based data analysis with machine learning technique was applied to fertilizer treatment of maize. We address this issue as our country depend on agricultural field rather than others. Maize has a bright future. To predict fertilizer treatment of maize dataset was comprised of ground coverage region which highlights the green pixels of a maize image. For calculating green pixels from an image we used "Can Eye" tool. The achievement of machine learning approaches is highly dependent on quality and quantity of the dataset which is used for training the machine for better classification result. For this perseverance, we have collected images from the maize field directly. Then processed those images and classified the data into four classes (Less Nitrogen=-N, Less Phosphorus=-P, Less Potassium=-K and NPK) to train our machine using decision tree algorithm to predict fertilizer treatment. We have got 93% classification accuracy for decision tree. Finally, the outcome of this paper is fertilizer treatment of a maize field based on the ground cover percentage, and we implemented this whole paper work using an android platform because of the availability of android mobile phone throughout the world.

Corresponding Author:

Nusrat Jahan,
Department of Computer Science and Engineering,
Daffodil International University,
Dhaka-1207, Bangladesh.
Email: nusratjahan.cse@diu.edu.bd

1. INTRODUCTION

Agriculture sector is backbone for every developing country like Bangladesh. Rice and jute are considered as the primary crops while maize and wheat are attracting greater importance nowadays. It is very common in our country that most of the crops suffer from different diseases due to lack of different fertilizer contained in the field. For this reason, we got interest to work on fertilizer treatmentof different crops. We started with maize to provide proper fertilizer treatment after leaves image analysis. Ground coverage collectively used in agronomic studies that can be used in smart agriculture applications to perform automatic treatment of crops. Ground cover analysis including estimation (the percentage of an area covered by vegetation), classification (instance-based labeling of vegetation), and segmentation (demarcation and extraction of regions of vegetation), is most often done by using image based methods, that discussed by F. Baret and G. Guyot, 1991 [1]. Especially, our study will be helpful for farms located in third world countries that have little access to high technology machinery but need to produce a lot of crops. Digital image analysis technique is considered to perform ground cover percentage and decision tree algorithm to classify the maize dataset. The ground cover percentage was considered in our method to make a decision about the health of crop because the prediction is based on the fertilizer treatment as fertilizers enhance the growth of crops.

This paper investigates the use of machine learning for fertilizer treatment of maize. Figure 1 depicted the process of machine learning approach that we followed to predict fertilizer treatment with the help of our proposed application. We divided the whole process into four levels – Input, Preparation, Classification, and Output.
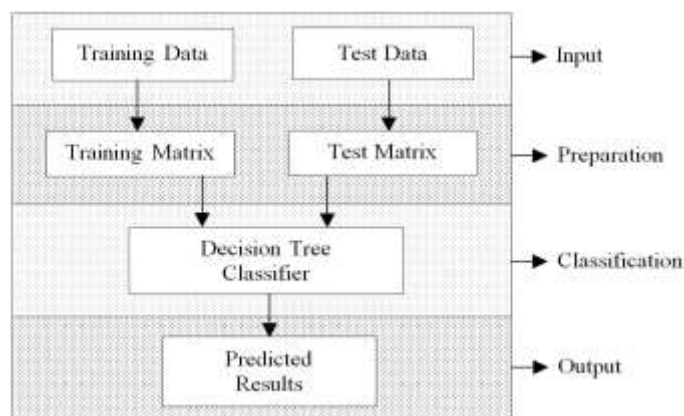

Figure 1. ML steps for fertilizer treatment of maize

In 2018 and after that in 2020 proposed an approach to measure Nutrient Deficiency in Crops. For producing food as much as needed nutrient ratio is crucial. In these studies they applied CNN to analysis plants leaves to calculate fertilizer deficiency [2, 3]. Von Ryan P. Marcelo and Joe G. Lagarteja in 2020 worked on identifying nutrient deficiency through image processing and achieved 91.5% accuracy [4]. Another research ended on soil nutrient analysis. Here, Phosphorus (P), Potassium (K), Organic Carbon (OC), Boron (B) as well as soil pH level considered to ensure crop productivity. Extreme Learning Machine (ELM) was proposed to solve these 5 nutrient classification problems. Finally, for four nutrients out of five got 80% accuracy to calculate proportion of those elements in soli. Stephen M. Ichami et al. also worked on fertilizer and nutrient efficiency but investigated the results only from African [5, 6]. In 2019 four machine learning algorithms was applied to measure maize yield and nitrate loss prediction and in 2020 Sudarshan Dutta also worked with maize for smallholder farms [7, 8]. Automated farming prediction was another work to provide entire farming system through machine learning [9].

Abd El-Aziz et al. and Cao et al. noted that fertilizers are main ways of plant growth which is added with the soil to improve natural growth. Where NPK is the primarily composed of three main elements- Nitrogen, Phosphorus, and Potassium, each of these are very essential in plant development. The estimation of ground cover proportion can be used in smart agriculture applications to perform effective treatment of crops [10, 11]. In Table 1, the three main macronutrients and their functions which are considered essential elements in plant survival and growth [12].

Table 1. The macronutrients for crops

Macronutrients	Function
Nitrogen(N)	Ecessary for Leaf growth
Phosphorus(P)	Growth of roots, flowers, seeds and fruit
Potassium(K)	Strong stem growth, movement of water in plants, promotion of flowering and fruiting

Furthermore, we studied some works where researchers performed image analysis on crops. Leaf disease measurement by using segmentation and histogram threshold value was noted by Sanjay B. Patil and Shrikant K. Bodhe [13]. Affected leaf area calculation by using number of pixel value, in this case authors used binary image arrangement [14]. Sanjay B. Patil in 2011 described the process to measure leaf area of sugarcane with the help of pixel number statistics, where authors used an object as a reference to measure leaf area [15]. Ground cover analysis and crop segmentation has been an active field of research over decades. Researchers applied image processing skills to measure quick and accurate leaf area in different purposes. Piyush Chaudhary described a procedure to calculate the leaf area with the help of

CIELAB color space. Here, threshold value calculated using OTUS's method, and leaf area was claculated by the following five phases- i. image attainment, ii. image preprocessing, iii. leaf area segmentation, iv. area filling and v. leaf area calculation. Accuracy of this method was 99% [16]. Some researchers have used contour extraction technique described by Tan Feng, Wang Chun. They used it for leaf area segmentation [17].

We also found some other papers those focused on applying machine learning algorithm in crop health analysis. Karishma. M et al. reviewed some research papers where they found SVM and ANN are mostly used algorithms for machine learning. In this paper they enhanced the growth rate of crop the artificial intelligence used for implementing the significant modern technology [18]. H. Al-Hiary et al. described the procedure to detect and classify the leaf disease where they applied K-means clustering and also used Neural Networks for recognition [19]. In 2013, Sonali Jain worked with SVM algorithm for feature extraction for the content based image [20]. The Weka(3.8.0) tool provides many algorithms, such as the Random Forest, J48, JRip and Support Vector Machine (SVM) [21]. Here, we also used this tool for exploring train data using machine learning algorithm. It is easy to apply any algorithm and also we can control some attributes of an algorithm.

This is because, we combined ground cover analysis with decision tree approach to implement an application to predict maize fertilizer treatment from a maize image. Consequently, decision tree provided 93% accuracy. We arranged this paper as follows- initially we demonstrated our methodology to predict fertilizer treatment in Section 2. Ssection 3 for result discussion and implementation of proposed method was illustrated in Section 4. Finally, conclusion of our work with some findings discussed in Section 5.

2. RESEARCH METHODOLOGY

This section contains the description of our proposed work for maize fertilizer treatment. As we mentioned before, this methodology focused on image analysis procedure, decision tree algorithm for machine learning and ground coverage percentage for fertilizer treatment.

2.1. Data collection and Image analysis

We considered 2 by 2 feet fixed size image frame for capturing and collecting image. This image should be collected from the top view of the maize crop. Next, we have to consider the fixed size images to predict the crop health conditions. Considering the number of maize crop's leaf, we classified the whole dataset into two categories- 3 leaves, and 6 leaves. Maize grows such a way that it would have 3 or 6 or 10 leaves in its lifetime. In Figure 2, gives the demonstration of sample images with 3 and 6 leaves in predicting the fertilizer treatment. After that, we used an image analysis tool ("Can Eye") to measure green pixel (Ground coverage percentage) of an image to prepare our dataset, as like Table 2. We collected the raw data (fixed size image) in total 245 instances from the International Maize and Wheat Improvement Centre (CIMMYT), Bangladesh from their maize research field.

However, users have to remember that they need to take a fixed size image from the top of the crops. Different devices capture images using different camera resolution, though we did not provide any functionality to trim the images. For this reason, at first we converted the image into a fixed resolution image. Then the following (1) is followed for calculating the ground cover percentage of a specific image:

$$Ground\ Cover\ \% = \frac{Total\ Green\ Pixel}{Total\ Pixel} * 100 \tag{1}$$

After calculating ground coverage percentage from each image, our next step was to classify dataset into four classes. Table 2 presented the sample dataset that we considered to implement the application using decision tree algorithm.

Table 2. Sample data

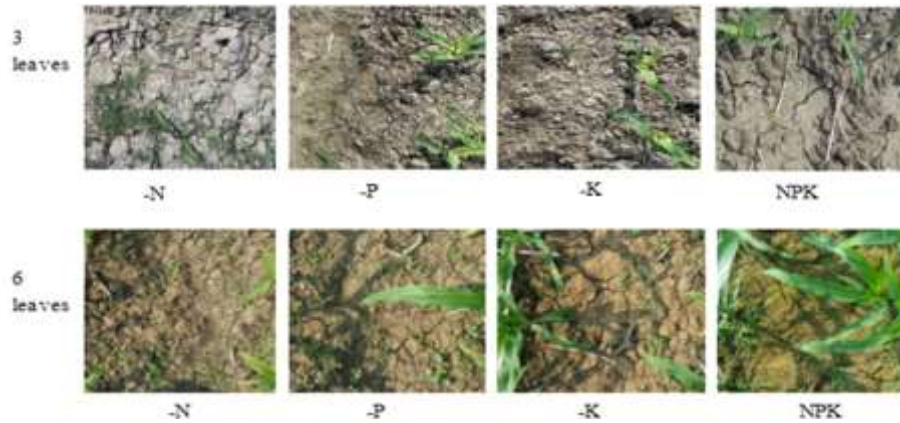| No. of Leaf	Ground_Coverage %	Fertilizer	Describe Fertilizer
3	4.27	NPK	Contain three components (N, P, K)
3	2.88	lessK	Lack of Potassium
3	2.76	lessP	Lack of Phosphorus
3	1.8	lessN	Lack of Nitrogen
6	42.59	NPK	Contain three components(N, P, K)
6	14.81	lessK	Lack of Potassium
6	10.52	lessP	Lack of Phosphorus
6	6.81	lessN	Lack of Nitrogen

Figure 2. Sample maize image

2.2. Decision tree

Decision Tree (DT) is one of the most popular supervised machine learning approach where the data is continuously divided into groups according to particular parameters. The tree can be explained by two main entities, namely decision nodes and leaves. This algorithm use "Information Gain" (Equation 3) approach to select the appropriate root node. We choose larger Information Gain attribute as a root node. When calculate Information Gain, it follow another property called "Entropy" (2) [22].

Decision tree is an easy and simple algorithm as well as implementation is not so complex that is why we chosen this algorithm. Decision tree algorithm works better on these types of dataset. Figure 3 to exhibit the decision tree algorithm that is based on ground cover percentage to predict the fertilizer treatment of maize.

```
Ground_Cover <= 2.87
| Ground_Cover <= 2.16: lessN (35.0)
| Ground_Cover > 2.16: lessP (32.0/2.0)
Ground_Cover > 2.87
| Ground_Cover <= 3.44
|| Ground_Cover <= 2.91: lessP (7.0/3.0)
|| Ground_Cover > 2.91: lessK (27.0)
| Ground_Cover > 3.44
|| Ground_Cover <= 8.65
||| Leaf_No <= 3: NPK (40.0/3.0)
||| Leaf_No > 3: lessN (29.0)
|| Ground_Cover > 8.65
||| Ground_Cover <= 15.18
|||| Ground_Cover <= 12.19: lessP (27.0/1.0)
|||| Ground_Cover > 12.19: lessK (21.0)
||| Ground_Cover > 15.18: NPK (27.0)
```

Figure 3. Decision tree steps for predicting fertilizer treatment of maize

$$Entropy(S) = \sum_{i=1}^{n}(-P_i log_2 P_i) \tag{2}$$

$$Information\ Gain(S,x) = Entropy\ (S) - \sum_{v=x} \frac{|S_v|}{S} + Entrpoy\ (S_v) \tag{3}$$

3. RESULTS AND ANALYSIS

We considered confusion matrix and other parameters to analysis the results of our proposed approach [23]. Confusion matrix is an approach that is used for presenting the performance of an algorithm. In confusion matrix there are four parameters: True positives (TP) provided the true result, True negatives (TN) provided the false result, False positives (FP) provided true result but actually this result is false.

Finally, False negatives (FN) provided false result but actually the result is true [24]. Table 3 represents the confusion matrix with four parameters. Table 4 for presenting the confusion matrix of decision tree. There are actual 68 lessN values but our method can correctly predict 64 data as a lessN class and incorrectly predict 4 data as a lessP class. To present our proposed approach we used confusion matrix, in this regard we prepared Table 4. It is going to describe the classification results of our model, where there were four classes

Table 3. Definition of confusion matrix

	Predict True	Predict False
Actual True	TP	FN
Actual No	FP	TN

Table 4. Confusion matrix of proposed model

Class	lessN	lessP	lessK	NPK
lessN	64	4	0	0
lessP	0	55	5	0
lessK	0	4	47	4
NPK	0	0	1	63

Next, we presented the results of decision tree algotirhm with six major parameters- True Positive Rate (TP Rate), False Positive Rate (FP Rate), Precision, Recall, F-Measure and Receiver Operating Characteristic curve or ROC Area [25]. To present all these parameters with their value we prepared Table 5. Here, we individually demonstrsted each class with classification value.

Table 5. The value for six measurement factors

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
lessN	0.939	0	1	0.939	0.969	0.984
lessP	0.917	0.043	0.873	0.917	0.894	0.971
lessK	0.855	0.032	0.887	0.855	0.87	0.944
NPK	0.984	0.022	0.94	0.984	0.962	0.979
Weighted Avg.	0.927	0.023	0.928	0.927	0.927	0.97

In essence, it showed that our implemented application of our model will be produced 93% accurate result. In the next section, we will discuss on implementation work.

4. SYSTEM IMPLEMENTATION

Now, we will describe the implementation of the proposed method to predict the fertilizer treatment of maize. Our aim was to help farmers with this application to produce maize successfully. As we know, there are lakes of agriculturalist in rural area, so it is crucial to ensure crop's growth rate as much as possible. In addition, people are more familiar with smart phone now a days and android is the common platform. That is wahy we also pick android platform to implement our approach.

Figure 4, illustrated the flow chat of our whole work. It is clear from the flow chart that we worked with an image after that, calculated the ground cover percentages. Next, we compared new ggound coverage value with threshold value to find the best fertilizer treatment class.

Initially, a user will have to take an image of maize from the crop field. In the next step, need to select the number of leaves contained in that image as we mentioned in previous section. The initial interface of this app is illustrated in Figure 5 "(a) First interface". The first interface provided a survice, where a user/farmer will have to select an option to indicate the number of leaves containe in that maize image. Though, we could not able to collect any 10 leave image. In the next interface user must select the captured image to perform the image analysis part which is depicted in Figure 5(b).

Here, a user will get "CHOOSE IMAGE" option to go for the next interface. Two options available here – one for capturing image directly using the camera and another for choosing image from phone gallery. After capturing image our will device perform an image analysis process to calulete ground coverage persentage. Decision tree algorithm will be used here to determine the fertilizer treatment of this particular maize crop. Finally, the interface to display the result is presented in Figure 6(a). If any image with lack of fertilizer then our application will detect it to provide a smooth maize farmimg system.
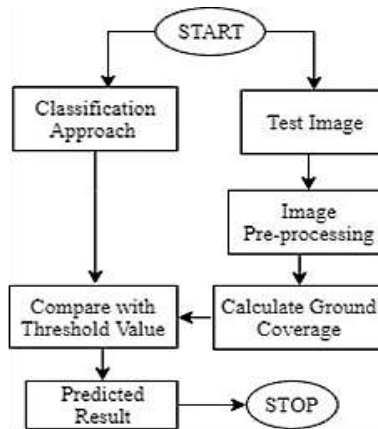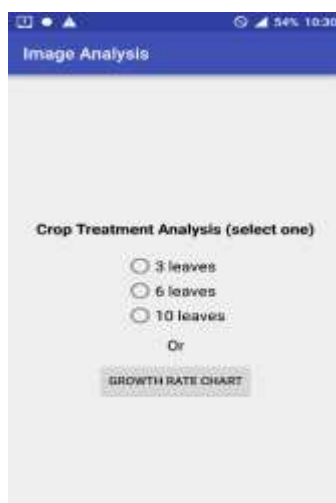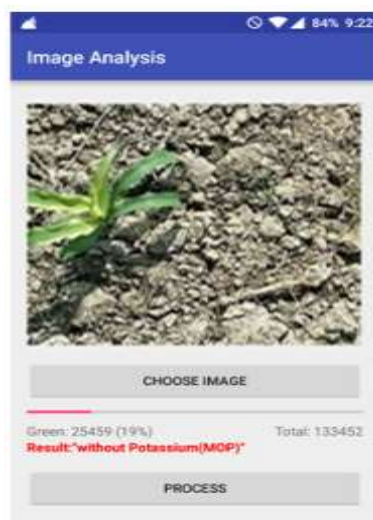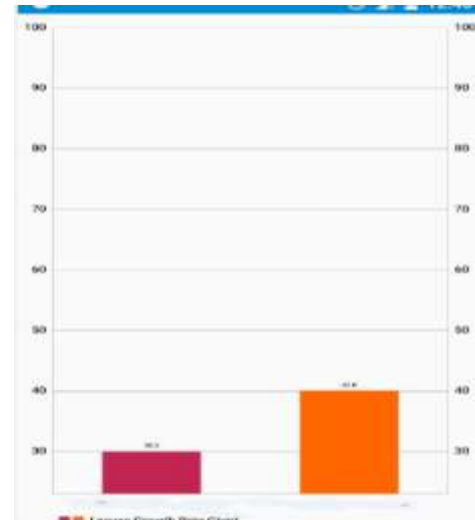
Figure 4. Flow chart for deployment

| (a) | (b) |

Figure 5. (a). First interface (b). Designed for image acquisition

| (a) | (b) |

Figure 6. (a). Ground coverage calculation, (b). Growth rate chart

Here, a user also will be able to determine the crop growth rate by camparing green pixel percentage of multiple maize images. If anyone wants to know about the growth condition of the crop then the user can click the growth rate button, which depict in Figure 6(b). To run this interface user need to capture more than one image, after that system will be able to produce the growth rate chart to understand the growth condition of the crop.

5. CONCLUSION

In this paper, we evaluated the effectiveness and applicability of decision tree in the domain of maize fertilizer treatment. In this study, the outcome of decision tree algorithm is considerable for predicting maize fertilizer treatment. However, for data classification purpose we used machine learning tool Weka 3.8.0. We believe this paper work would be benefited for farmers to maize fertilization treatment. Maize production can be the way to solve food crisis in future. In future, our desire to collect more accurate data from different crops field to perform irrigation treatment which will be incorporating with this paper work and to maintain acceptable level of accuracy deep learning algorithm will be considered. In order to achieve good accuracy level for crop treatment, we will have a plan to collect more accurate spatiotemporal data. These types of study are actually necessary to solve population based food crisis problem. Maize can be a major crop in many countries with proper care.

REFERENCES

[1]   F. Baret and G. Guyot, "Potentials and limits of vegetation indices for LAI and APAR assessment," *Remote sensing of environment*, vol. 35, no. 2, pp. 161-173, 1991.
[2]   Amirtha T, Gokulalakshmi T, Umamaheswari P, and T Rajasekar M. Tech, "Machine Learning Based Nutrient Deficiency Detection in Crops," *International Journal of Recent Technology and Engineering (IJRTE)*, ISSN: 2277-3878, Vol. 8, no. 6, March, 2020.
[3]   U. Watchareeruetai, *et al.*, "Identification of Plant Nutrient Deficiencies Using Convolutional Neural Networks," *2018 International Electrical Engineering Congress (iEECON)*, Krabi, Thailand, pp. 1-4, 2018.
[4]   Von Ryan P. Marcelo, Joe G. Lagarteja, "Corzea: Portable Maize (Zea Mays L.) Nutrient Deficiency Identifier," *International Journal Of Scientific & Technology Research*, vol. 9, no. 2, February 2020.
[5]   Suchithra, M. S. and Pai, M. L., "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters," *Information Processing in Agriculture*. doi:10.1016/j.inpa.2019.
[6]   Ichami, S.M., *et al*., "Fertilizer response and nitrogen use efficiency in African smallholder maize farms," *Nutrient Cycling in Agroecosystems*, vol. 113, pp. 1-19, 2019.
[7]   Mohsen Shahhosseini , Rafael A Martinez-Feria , Guiping Hu and Sotirios VArchontoulis, "Maize yield and nitrate loss prediction with machine learning algorithms," *Environmental Research Letters*, Vol. 14, no. 12, 4 December 2019.
[8]   Dutta, S., Chakraborty, S., Goswami, R., Banerjee, H., Majumdar, K., Li, B. and Jat, M. L., "Maize yield in smallholder agriculture system-An approach integrating socio-economic and crop management factors," *PloS one*, vol. 15, no. 2, e0229100, 2020.
[9]   Siddique, T., Barua, D., Ferdous, Z. and Chakrabarty, A., "Automated farming prediction," *Intelligent Systems Conference (IntelliSys)*. doi:10.1109/intellisys.2017.8324214, 2017.
[10] A. El-Aziz, Nahed G., "Stimulatory effect of NPK fertilizer and benzyl adenine on growth and chemical constituents of Codiaeum variegatum L. plant," *Am.-Eurasian J. Agric. Envi. Sci*., vol. 2, pp. 711-719, 2007.
[11] Q. Cao, *et al.*, "Developing a new crop circle active canopy sensor-based precision nitrogen management strategy for winter wheat in North China Plain," *Prec. Agric*, vol. 18, pp. 2-18, 2017.
[12] Fertilizer. [Online]. Available: https://en.wikipedia.org/wiki/Fertilizer. Access date: 12/01/2018.
[13] S. B. Patil and S. K. Bodhe, "Leaf Disease Severity Measurement using Image Processing," *International Journal of Engineering and Technology*, vol. 3, no. 5, 2011.
[14] H. D. Marathe and P. N. Kothe, "Leaf Disease Detection Using Image Processing Techniques," *International Journal of Engineering Research & Technology (IJERT)*, Vol. 2, no. 3, March, 2013.
[15] S. B. Patil, "Image Processing Method to Measure Sugarcane Leaf Area," *International Journal of Engineering Science and Technology*, ISSN: 0975-5462 Vol. 3 No. 8 August, 2011.
[16] P. Chaudhary, S. Godara, A. N. Cheeran and K. Chaudhar, "Fast and Accurate Method for Leaf Area Measurement," *International Journal of Computer Applications*, (0975-8887), Vol. 49, no. 9, July, 2012.
[17] T. Feng and W. Chun, "Calculating the Leaf-Area Based on Non-Loss Correction Algorithm," *International Conference of Information Science and Management Engineering, IEEE*, pp. 75-78, 2010.

[18] K. Mohiuddin and M. M. Alam. "A Short Review on Agriculture Based on Machine Learning and Image Processing," *Acta Scientific Agriculture* vol. 3, no. 5, pp. 55-59, 2019.

[19] H. Al-Hiary, *et al.*, "Fast and Accurate Detection and Classification of Plant diseases," *International Journal of Computer Applications*, (0975-8887), Vol. 17, no. 1, March, 2011.

[20] S. Jain, "A Machine Learning Approach: SVM for Image Classification in CBIR," *International Journal of Application or Innovation in Engineering and Management*, (2319-4847), vol. 2, no. 4, April, 2013.

[21] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.

[22] A. Navada, A. N. Ansari, S. Patil and B. A. Sonkamble, "Overview of use of decision tree algorithms in machine learning," *2011 IEEE Control and System Graduate Research Colloquium*, Shah Alam, pp. 37-42, 2011.

[23] Ting K.M., "Confusion Matrix," *In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer*, Boston, MA, 2017.

[24] Simple guide to confusion matrix terminology. [Online]. Available: http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/. Access date: 15/02/2018.

[25] Majnik, Matjaz and Bosnic, Zoran., "ROC Analysis of Classifiers in Machine Learning: A Survey," *Intelligent Data Analysis*, vol. 17, no. 3, pp. 531-558, 1 jan. 2013.

BIOGRAPHIES OF AUTHORS

Nusrat Jahan was born in Barisal, Bangladesh. She completed her Graduation in Computer Science and Engineering from Patuakhali Science and Technology University, Bangladesh and Master in Information Technology from Institute of Information Technology, University of Dhaka. She has 5 years of teaching experience. She is currently working as a Senior Lecturer in computer science and engineering department at Daffodil International Unoiversity. Her research interests are machine learning, data Analysis and image processing.

Rezvi Shahariar is an Assistant Professor in the Institute of Information Technology, University of Dhaka. He completed Bachelor and Master Degrees in Computer Science and Engineering from the University of Dhaka. His research interests include Ad hoc Networks, Sensor Networks, and Security Issues in VANETs, Machine Learning, NLP and Software Engineering.