International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 1052-1058 www.ijaem.net ISSN: 2395-5252

Application of Machine Learning Techniques for Extraction of Soil Features for Pattern Classification

Srigowri M V[1*], Dr.BhavanishankarK[2],Soujanya R Bhat[1], Sampoorna K[1]
[1]*Alumni of Dept of CSE, RNSIT, Bengaluru, India*
[2]*Associate Professor, Dept of CSE, RNSIT, Bengaluru, India,*
[*]*Corresponding Author: SRIGOWRI M. V*

Date of Submission: 30-06-2020 Date of Acceptance: 15-07-2020

ABSTRACT: Agricultural researchers all over the world insist on the need for an efficient mechanism to predict and improve the crop growth. The need for a crop prediction system with accurate recommendation is highly felt among farming community. The complexity of predicting the most suitable crop is high due to multi-dimensional variable metrics and unavailability of predictive modelling approach, which leads to loss in crop yield.In this research work, a general architecture for crop prediction using soil images and the appropriate region as inputs to predict the most suitable crop for cultivation is developed. Training data set for finding the type and texture of soil were collected by capturing the soil images using a high resolution camera in the University of Agricultural Sciences, Bengaluru. The experimental procedures for soil analysis used in this University are observed and studied and the results are compared with the proposed system for validating the results. In the proposed system, soil images are analyzed and processed using image processing techniques. Machine learning and deep learning techniques are used for feature extraction. Data mining techniques are applied on the features extracted for predicting the crop.The trained classifier predicts the name of the crop suitable for the soil. In this work, color of soil, pH value, the soil type and texture have been successfully determined for input soil images. All the algorithms are coded in Python language. This system takes less computational time when compared to the manual experimental procedures. The results obtained are discussed with a domain expert from the Agricultural University for validation purpose. The results obtained are found to be satisfactory with an overall accuracy of 80%.

Keywords: Agriculture,Decision Tree Classifier, Deep convolutional neural network, Python, Soil Images, Crop prediction

I. INTRODUCTION

Agriculture is an important occupation in the world and it requires immense agricultural resources to meet the needs of people. The significant agricultural products include coffee, wheat, rice, maize, ragi, sugarcane, cotton, etc. Several agricultural techniques are required that allows a rational use of inputs, at variable rates, and also allows a better understanding of the variability in soil and soil features.

The characteristics of soil play a big part in the plant's ability to extract water and nutrients. If plants are to grow to their potential, the soil must provide a satisfactory environment for the growth. Plants obtain most of their oxygen and carbon from the air by photosynthesis and hydrogen is obtained, directly or indirectly, from the water in the soil. These three elements together make up over 90 percent of fresh plant tissue. However, plants cannot survive without essential nutrients that they obtain from the soil, such as nitrogen, phosphorus, potassium, calcium, magnesium and sulphur [25].

Agricultural researchers over the world insist on the need for an efficient mechanism to predict suitable crops based on soil types, soil features and meteorological information. The traditional methods involve sieve analysis for coarse-grained soils and hydrometer analysis for fine grained soils. On repeated use, the sieve openings get distorted and give erroneous results. These methods are time consuming and redundant. Soil test results generally take a couple of weeks to arrive back from the lab. More importantly, sieve analysis does not necessarily measure the particle diameter in the conventional true sense. Soil particles are three-dimensional and particle size based on sieving captures the intermediate dimension.

The use of computers not only automates the process of soil characterization, but also makes it more objective. The chances of human error are minimized, saving a lot of energy, time and most importantly, money. In soil science, image analysis can be used to measure specific features of the soil or vegetation. Both characteristics of the crops and the soil material itself can be measured using image analysis. Images are generally analyzed to determine either color of an object or to find and analyze shapes and patterns. There are many more types of analyses that can be done to extract useful information from images related to soil science. Some examples are Fourier transforms, fractal dimension analysis, texture analysis etc. In this work, soil images are used as inputs which will be analyzed and processed using image processing techniques. Machine learning and deep learning techniques are used for feature extraction and classification for crop prediction in an appropriate region based on the soil type.

The paper is organized as follows: Section I contains the introduction of soil features and need of soil image analysis for crop prediction. Section II contains the related work of soil image analysis and machine learning techniques used. Section III contains the methodologies used for determining soil color, pH value of soil, soil texture and prediction technique. Section IV contains the algorithms, results and discussion. Section V concludes the research work and future scope.

II. RELATED WORK

To carry out this research work, several journals, conference papers and text books were studied. They are briefly described below.

In [1], the authors have described a method to find the soil pH value using a digital image processing technique. Soil pH property is used to describe the degree of acidity or basicity which affect nutrient availability and ultimately plant growth.Due to concentration of organic matters, presence of water and oxidation are the influenced factors of pH and color association. Correlation between digital value and soil pH values should be helpful in determination of soil pH value for different type of soils.

The authors Mrutyunjaya R D et. al [11] discuss the process used to develop their engineering capstone project “Soil moisture Assessment Software”. Moisture content in soil is one of the main component which plays an important role in yield of crops. Image of the soil with different moisture content are captured and pre-processed to remove the noise in source image. The color and texture characteristics of moist soil are extracted. Color characteristics are analyzed using the RGB and the HSI models.

The author Magnus Persson [7] has made use of dye traces for image analysis. Dye traces have been used for many years by soil scientists investigating the effects of soil heterogeneity as they allow visualization of spatial flow patterns. Traditionally, image analysis of the dye photographs has only involved separation between stained and non-stained soil. The color as expressed in a specific color space of the corrected images can be related to dye concentration. The relationship between soil color and dye concentration is soil specific. The same fundamental principle of color consistency can also be used for determining other physical characteristics of soil material like organic matter content or water content.

The authors Pravat Kumar et. al [12], make use of image processing techniques that are applied to perform surface soil crack analysis. The geometric features of cracks, such as width, length, and surface area are estimated. These parameters are important, because they influence both the soil hydraulics and mechanics.

In the referred paper [15],a fuzzy classification methodhas been implemented for determining land suitability from soil profile observations and topography. Because conventional Boolean retrieval of soil survey data and logical models for assessing land suitability treat both spatial units and attribute value ranges as exactly specifiable quantities, they ignore the continuous nature of soil and landscape variation and uncertainties in measurement which can result in the misclassification of sites that just fail to match strictly defined requirements. The results obtained are compared with conventional Boolean methods and are found to be better.In papers [10,12,13,15], various experiments have been done on soil types and textures for increasing crop yield.

In paper [8], the authors have given taxonomy of Deep Convolutional Neural Network (DCNN). DCNN techniques are also used in papers [2,3,4,5,8].

In paper [6], the authors have used a large DCNN to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes.

The authors of the referred paper[9] have developed a crop yield prediction model (CRY) based on an adaptive cluster approach over dynamically updated historical crop data set to predict the crop yield and to improve the decision making in precision agriculture. Bee hive modelling approach is used to analyze and classify the crop based on crop growth pattern.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 1052-1058 www.ijaem.net ISSN: 2395-5252

III. METHODOLOGY

In the twenty-first century, agricultural researchers over the world insist on the need for an efficient mechanism to predict suitable crops, that can use a significant increase in the amount of data and information, which comes from a broad number of different sources such as meteorological information, location based information, soil features, etc.The objective of this research work is to predict the most suitable crop based on the soil images and topographical information. Using soil images, the soil pH value, soil type and soil texture are determined. The goal is develop an inexpensive method and also to reduce the time taken for crop prediction.A crop prediction model is developed in this research work, which is capable of suggesting the most suitable crop for the given input soil image. The idea is to automate the crop prediction instead of using the experimental methods. The model overcomes the disadvantage of using traditional methods and do-it-yourself techniques that are time consuming, redundant and expensive. The images of different soil types available in the University of Agriculture Sciences, Bengaluru, were captured by the authors of this paper for collecting the training data sets. The experimental methods used in the University and their results were referred for verifying with the results of the proposed techniques.

A. Proposed Crop Prediction System

Figure 1 shows the system architecture of the proposed crop prediction system. The input module of the application accepts the soil image and the location. The soil image is given as the input to the feature extraction module. The feature extraction module has the following sub-modules:

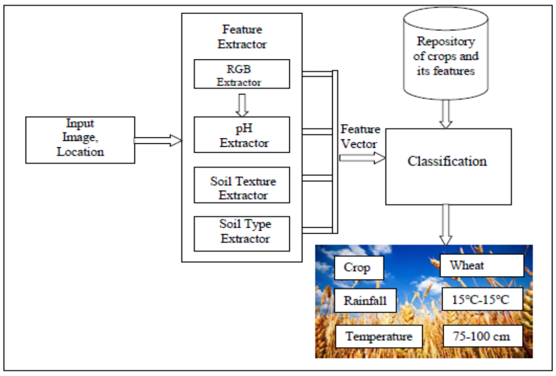

Figure 1: The System Architecture of the Crop Prediction System

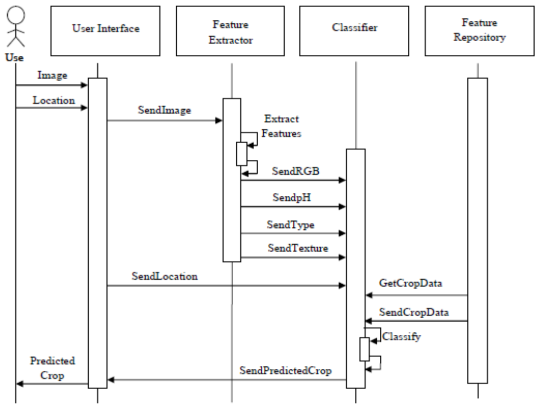

Figure 2: Sequence Diagram for the Crop Prediction System

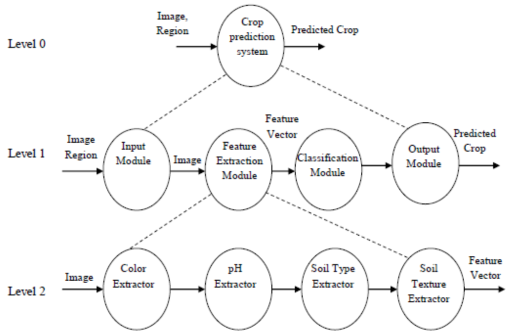

Figure 3: Data Flow Diagram for the Crop Prediction System

1. Color extraction: This sub-module extracts the most dominant color and RGB (Red, Green, Blue) values from the soil image.
2. pH determination: This sub-module makes use of the RGB valuesto determine the pH value of soil.
3. Soil Type determination: This sub-module makes use of a fully trained DCNN to classify the soil to one of the six soil types.
4. Soil Texture determination: This sub-module makes use of a fully trained deep convolution neural network to classify the soil to one of the twelve soil textures. The output of the module is a feature vector of the input soil image.

The sequence diagram and data flow diagram developed for the crop prediction system are shown in Figure 2 and Figure 3 respectively.

B. Soil Types and Soil Textures

Soil can be defined as the organic and inorganic material on the surface of the earth that provides the medium for plant growth. Soil develops slowly over time and is composed of many different materials. Inorganic materials, or those materials

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 1052-1058 www.ijaem.net ISSN: 2395-5252

that are not living, include weathered rocks and minerals.

Though there are more than ten soil types as per the literatures and internet resources, only 6 soil types physically available in the University of Agricultural Sciences used for their experiments are considered in this research work. They are: Alluvial Soil, Laterite Soil, Red Soil, Black Soil, Desert Soil and Mountain Soil.Figure 4 shows the sample images of red soil, laterite soil, black soil, alluvial soil respectively.

Soil Texture indicates the relative content of particles of various sizes, such as sand, silt and clay in the soil. Texture influences the ease with which soil can be worked, the amount of water and air it holds, and the rate at which water can enter and move through soil. The 12 soil textures used in this research work for classification are: Clay, Clay Loamy, Loam, Loamy Sand, Sand, Sandy Clay, Sandy Loam, Sandy Clay Loam, Silt, Silty Clay, Silty Clay Loam and Silty Loam. Figure 5 shows the sample images showing 4 different soil textures for the same soil type.

C. Algorithms developed

K-Means clustering algorithm [17] has been used in **Algorithm_1**named ColorExtraction, to find the Red, Green and Blue values in the given input soil image.

Algorithm_1: ColorExtraction
Input: Input soil image
Output: Color of the input soil image
Begin
Read every pixel of the soil image into a list #each pixel is made up of Red, green, blue components
form three clusters
Randomly choose three values from the list as centroids of each cluster
Repeat
for every pixel value in the list
find Euclidian distance to each centroid
Assign the pixel value to the nearest cluster
End for
Find new centroid for each cluster by finding the mean value
Until centroids do not change
Return the centroid of the cluster with most pixel values in it
#the most dominant color
End

Decision tree classifier that uses Classification and Regression tree algorithms has been used to determine the pH value of the soil using the R, G and B values determined using Algorithm_1.

A DCNN named Inception v3 model [19,20], as shown in Figure 6, has been used in this research work for determining the type and texture of the soil based on theimage. To predict the best crop for the input soil image along with the location, decision tree classifier has been used [16,17]. The decision tree has been trained initially using the training data sets.

All the algorithms were developed using Python programming language [18,21]. Suitable number of training data sets were collected and used as discussed in the next section. All the algorithms are not given in this paper because of space constraints.

Figure 4: Sample Images of Soil Types – Red Soil, Laterite Soil, Black Soil and Alluvial Soil, Respectively

Figure 5: Sample Images Showing Soil Textures

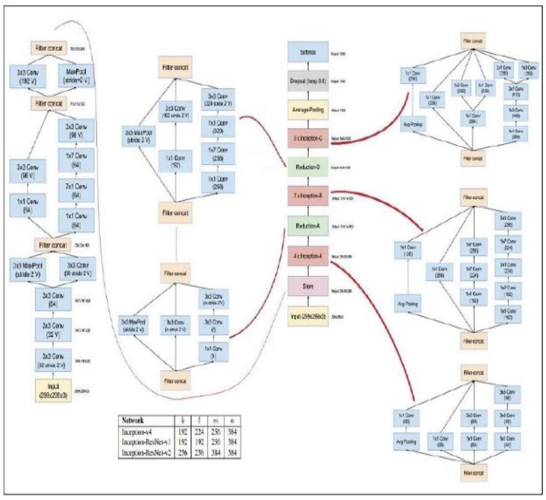

Figure 6: Inception V3 Model of DCNN

IV. RESULTS AND DISCUSSION

The results of the K-Means clustering algorithm usedto determine the R, G and B values and the Decision tree classifier (CART) used to determine the pH value of the soil using fifty data sets for which RGB values were determined are found to be satisfactory. Sample results of R, G and B values and the predicted pH values are shown in Table 1. These values were validated with the experimental results for those soil samples determined in the University of Agricultural Sciences.

Table 2 shows the results of all the algorithms used in this research work. A DCNN named Inception v3 model has been used in this work for determining the type and texture of soil for the input soil image. The algorithm is of transfer learning type. The training data set for determining the type of soil consisted of two hundred images in each of the six categories of soil type (total size is 1200). The training for finding the type of soil was done in five hundred iterations. The training data set for determining the texture of the soil included two hundred images in each of the twelve categories of soil (total size is 2400). The training for finding the texture of soil was done in five hundred iterations. While training two metrics - train accuracy and train loss are calculated on the go. These metrics show how well the network is doing on the data it is being trained on. Training accuracy usually keeps increasing throughout training. After every epoch, the model is tested against a validation set, (or test samples), and validation loss and accuracy are calculated. These numbers tell how good the model is at predicting outputs for inputs it has never seen before. Validation accuracy increases initially and drops as it over fit. The first two rows of results in Table 2 indicate the results of DCNN transfer learning algorithms used for Soil Type and Soil Texture. Accuracy of 90% and 80% are obtained for the test data sets for soil type and soil texture respectively.

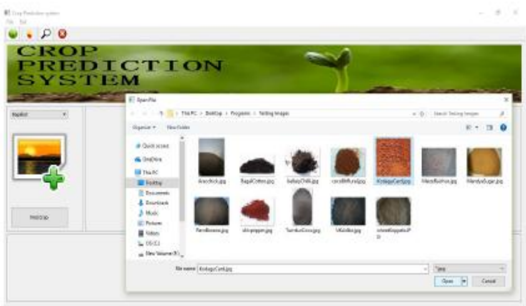

Figure 7: Snapshot of the user interface (GUI) developed for the crop prediction System

Figure 8: Snapshot showing the result for the suitable crop predicted - cardamom

Finally, to predict the crop for the input soil image, Decision Tree classifier (CART) has been used. The tree was trained on twenty data sets along with the location. The trained classifier predicts the name of the crop suitable for the soil and the classification accuracy is around 80%, as shown in the third row of results in Table 2.

The GUI for the application is developed using PyQt Python library modules, for making the crop prediction system more user friendly. Sample snapshots of the usage of the system and the result displayed during crop prediction are indicated in Figure 7 and Figure 8 respectively. This system takes less computational time when compared to the manual experimental procedures. The results obtained were discussed with a faculty in the Agricultural University for validation purpose. The results obtained are found to be satisfactory.

Table 1: Soil pH Value Predicted based on RGB values

Red	Green	Blue	pH Value
133	103	55	7.25
128	105	27	6.96
148	118	48	7.4
130	98	30	7.05
197	164	123	8.35

Table 2: Results of different algorithms used

Parameters Determined	Algorithm used	Training Data Set Size	Testing Data Set Size	Number of distinct classes	Classification Accuracy
Soil Type	Transfer learning	1200	20	6	90%
Soil Texture	Transfer Learning	2400	40	12	80%
Crop name	CART	20	10	-	80%
pH Value of soil	CART	50	10	-	Nearest pH Value was obtained for

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 1052-1058 www.ijaem.net ISSN: 2395-5252

					all the test cases
Color	K-Means clustering	-	10	-	RGB values of the soil were determined

V. CONCLUSION AND FUTURE SCOPE

In this research work, the soil images were captured using a high resolution camera in the University of Agricultural Sciences, Bengaluru, by the authors of this paper for data collection. Experimental studies were also observed and theoretical knowledge was gained by thorough discussion with the Professors in that university. The results obtained using the software algorithms developed by the authors were also discussed and validated.

The clarity in the soil images can be improved by using image pre-processing techniques. More number of soil types can be collected from other sources and information from more locations in India can be collected for future work. The results can be compared by developing other suitable classifiers.

ACKNOWLEDGMENT

The authors are grateful to the management and authorities of RNSIT, Bengaluru and VTU, Belgaum. The authors acknowledge the support and the guidance extended by the Professors of University of Agricultural Sciences,Bengaluru, to carry out this research work. This paper is based on the project work done in the final year B.E by the authors, under the guidance of Dr.Bhavanishankar K.

REFERENCES

[1]. Vinay Kumar, Rakesh Kumar, Binod Kumar Vimal, Mukesh Kumar and Rakesh Kumar, “Determination of soil pH by using digital image processing technique”, Journal of Applied and Natural Science, Volume 6, Issue 1, pp. 14-18, ISSN : 0974-941, January 2014.

[2]. Pratik Devikar, “Transfer Learning for Image Classification of various dog breeds”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 5, Issue 12, pp. 2707 – 2715, ISSN: 2278 – 1323, December 2016.

[3]. Rajat Raina, Alexis Battle, Honglak Lee, Benjamin Packer and Andrew Y. Ng, “Self-taught learning: transfer learning from unlabeled data”, Proceedings of the 24th International conference on Machine learning, pp.759-766, 20-24 June 2007.

[4]. S. J. Pan and Q. Yang, "A Survey on Transfer Learning", IEEE Transactions on Knowledge and Data Engineering, Volume 22, Issue 10, pp. 1345-1359, Oct. 2010.

[5]. C. Szegedy, "Going deeper with convolutions", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 1-9, 2015.

[6]. Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton “ImageNet Classification with Deep Convolutional Neural Networks”, Advances in neural information processing systems, pp. 1-9, 2012.

[7]. Magnus Persson, “Image analysis in soil science”, A review of different image analysis methods in the field of soil science, pp. 141-148, 2010.

[8]. Suraj Srinivas, Ravi Kiran Sarvadevabhatla, Konda Reddy Mopuri, Nikita Prabhu, Srinivas S, S Kruthiventi and R. Venkatesh Babu, “A Taxonomy of Deep Convolutional Neural Nets for Computer Vision”, Frontiers in Robotics and AI, Indian Institute of Science, Bangalore, arXiv:1601.06615v1 [cs.CV], 25 January 2016.

[9]. M. Gunasundari Ananthara, Dr. T. Arunkumar and Ms. R. Hemavathy, “CRY – An improved Crop Yield Prediction model using Bee Hive Clustering Approach for Agricultural data sets”, Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, ISSN: 978-1-4673-5845-3/13/$31.00, IEEE, 21-22 February 2013.

[10]. Wu Fan, Chen Chong, Guo Xiaoling and Yu Hua, “Prediction of crop yield using big data”, 8th International Symposium on Computational Intelligence and Design, ISSN: 978-1-4673-9587-8/15 $31.00, IEEE, 19th January 2015.

[11]. Mrutyunjaya R.,Toufiq A. Badebade, Megha M. Jain and Ashwini R. Maigur “Estimation of Moisture Content in Soil Using Image Processing”, International Journal Of Innovative Research & Development, ISSN 2278 – 0211, April 2014.

[12]. Pravat Kumar S, Gouri Sankar Bhunia and Ramkrishna Maiti, “Soil crack morphology analysis using image processing techniques”, Model. Earth Syst. Environ. (2015) 1:35, DOI:10.1007/s40808-015-0036-z, November 2015.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 1052-1058 www.ijaem.net ISSN: 2395-5252

[13]. Mohamad Ayob Mohamadi and Ataollah Kavian, "Effects of rainfall patterns on runoff and soil erosion in field plots ",International Soil and Water Conservation Research 3 (2015) 273–281, October 2015.
[14]. Andrej Karpathy, Li Fei-Fei, "Deep Visual-Semantic Alignments for Generating Image Descriptions", arXiv:1412.2306v2 [cs.CV], 14 Apr 2015.
[15]. M. D. Soriano-Soto, A. Calvo-Cases, C. Boix-Fayos and A. C. Imeson, "Effect of Climate on Some Soil Properties and Related Thresholds Controlling the Erosional Response of Soils in a Limestone Area ", Phys. Chem. Earth, Volume 20, No. 3-4, pp. 281-286, 1995.
[16]. Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing,3rd Edition, Pearson, ISBN-13: 978-0131687288, 2008.
[17]. Alpaydin, Ethem (2010). Introduction to Machine Learning. London: The MIT Press. p. 110. ISBN 978-0-262-01243-0. Retrieved 4 February 2017.
[18]. Jan Erik Solem, Programming Computer Vision with Python, O'Reilly, ISBN:978-1-4493-1654-9, March 2012.
[19]. Simon Haykin, Neural Networks and Learning Machines, 3rd Edition, Pearson, Prentice Hall, ISBN-10: 0131471392, 2009.
[20]. Michael A. Nielsen, Neural Networks and Deep Learning, Determination Press, 2015.
[21]. Zed A. Shaw, Learn Python the Hard Way, 3rd edition, Addison Wesley Publications, ISBN-13: 978-0321884916, October 2013.

Authors Profile

Srigowri M V is the corresponding author of this paper. She did her B.E. in CSE, RNSIT, Bengaluru. She has worked as a Software Engineer for Robert Bosch Engineering & Business Solutions Pvt Ltd, Bengaluru. She is joining for MTech program during Aug/Sept 2020

Dr. Bhavanishankar K is working as Assistant Professor in Department of CSE, RNSIT, Bengaluru from past 10 years. He has a total teaching experience of 17 years. He has published more than 10 research papers in reputed international journals and conferences.

Soujanya R Bhat did her B.E. in CSE, RNSIT, Bengaluru. She has worked as a Software Engineer forAurigo Software Technologies, Bengaluru. She is doing her MS program in Arizona State University

Sampoorna K did her B.E. in CSE, RNSIT, Bengaluru.She did her MS program in University of Texas, Dallas. She is presently working in Amazon, USA.

International Journal of Advances in Engineering and Management
ISSN: 2395-5252

IJAEM

Volume: 02 Issue: 01 DOI: 10.35629/5252