International Journal of Intelligent Computing and Technology
IJICT
ISSN: 2457-0249 (Online)

Website: https://www.ijict.com **Jan (2022) - Vol.5, Iss.2 Pages:63-73**

A STUDY ON INTEGRATED K MEAN ALGORITHM USING IMAGE CLASSIFICATION

R Shalini[1], C Pradeepa[2], A Emima[3]
[1,2]B.Sc., [3]Assistant Professor, PG Department of Computer Science
Holy Cross College, Trichy, Tamilnadu

Article History- Received: June 2021; Published: Jan 2022

Abstract

Agricultural productivity is a number of issues that are highly dependent on the economy. This is one of the reasons why the detection of diseases in plants plays a crucial role in the field of agriculture, as it is quite common to have diseases in plants. If proper care is not taken in this area, it will have serious effects on plants and on the quality, quantity or commodity concerned. For example, a disease called a small plant disease can be a dangerous disease found in pine trees in us. Detection of disease by some automated technique is helpful because it decreases the large-scale monitoring function in large-scale crop farms, and at a very early stage it detects symptoms of disease, i.e. When they appear on the leaves of a plant. This paper introduces an image segmentation technique algorithm which is used to automatically detect and identify plant-leaf diseases. It also includes studies on various classification methods of diseases that will be used for the identification of plant disease. Image classification refers to a computer vision method that is capable of classifying an image according to its visual content.

Keywords: *Image processing, Genetic algorithm, Plant disease detection Classification*

INTRODUCTION

In today's world, the agricultural land mass is quite literally a source of feed. The Indian economy is highly dependent on productive agriculture. Therefore, the identification of disease in plants plays a crucial role in the field of agriculture. Use of automated disease detection technique is useful for detecting a disease in a very initial stage. For instance, a disease called small plant disease may be a dangerous disease found in pine trees within us. The tree affected exhibits stunted growth and dies within 6 years. The effect is seen in Southern US areas of Alabama, Georgia. Early detection could be beneficial in such scenarios. The current procedure for disease detection is essentially expert naked eye observation through which plant disease identification and detection is achieved. To do so, a large team of experts is also needed as continuous plant monitoring, which will cost very high once we do it with large farms. In some nations, farmers do not have proper facilities or perhaps idea that they will contact experts at an equivalent time. As a result, consulting experts often cost a great deal of time as well. In such circumstances the suggested technique proves effective in tracking large crop fields. Automatic disease identification by seeing only the signs on the leaves of the plant often makes it as simple as it is. This also supports machine vision in providing image-based automated process control, inspection, and visual recognition of robot guidance disease is more difficult and at the same time less effective, and can only be achieved in limited areas.

Whereas it will take less effort, less time and become more effective if automatic detection technique is used. Some general conditions seen in plants are brown and yellow stains, early and late scorch, et al. fungal, viral, and bacterial diseases. Picture analysis is used to assess the affected disease region and to determine the color difference of the affected area [1,2,5]. Picture segmentation is the mechanism by which a picture is divided or grouped into different parts. There are currently several different ways to perform segmentation of images, ranging from the simple thresholding approach to advanced methods for segmentation of images in color. These pieces usually correspond to something that can be easily distinguished by humans and seen as individual objects. Computers have no means of intelligently recognizing objects, then many different methods of segmenting images are created. The process of segmentation is predicated on different features found in the image. This could be details on color, borders or image section [7,9]. We use the Genetic Algorithm for segmentation of color images. Reichenberg first implemented evolutionary computing within the 1960's. Other researchers then adopted his idea. Sometimes the evolutionary changes are small and seem insignificant at first glance, but in natural selection they play a neighborhood and thus the species survival.

LITERATURE SURVEY

In the paper [1], the author presented the various plant diseases as a major threat to the agricultural sector by reducing plant life. The aim of the present work is to develop a simple disease detection system for plant diseases.

Automatic detection of plant diseases was presented in paper [2]. It may prove beneficial in the monitoring of large crop fields and thus automatically detect diseases from the symptoms that appear on the leaves of the plant. Al-Bashish et al. (2011) developed a fast and accurate method in which the leaf diseases are detected and classified using k- means based

segmentation and neural networks-based Classification of the. Automatic classification of leaf diseases is based on high resolution multispectral and stereo images (Bauer et al., 2011).

In paper [3], the author presents fundamental approaches to the segmentation of images in digital image processing. Image Segmentation is a method used to separate object of interest from context. And he provided the methods for the classification of the image.

The author included the different types of plant disease in paper [4] and made a comparison table. The study also includes the different classification methods for diseases that will be used for the identification of plant diseases.

The author uses the k-mean algorithm for clustering a picture in paper [5]. His proposal was then brought forward by other researchers.

Kim et al. (2009) identified grape fruit peel diseases using color texture analysis. The texture characteristics are determined from the Spatial Grey-level Dependence Matrices (SGDM) and the classification is performed using a squared distance technique. Grape fruit peel may have been contaminated with a number of diseases, such as canker, copper burn, fat spot.

In paper [6] the main objective of this paper is to provide a segmentation algorithm which will be used in image classification tool in an effective manner. K-mean clustering algorithm with the edge's detection pixels in the feature space, each of them defined by its center. The logical representation of K-Mean algorithm.

METHODOLOGY

Machine learning

Machine learning is an artificial intelligence (AL) system that provides the ability to automatically learn and improve on knowledge without being programmed for exile. Machine learning focuses on the case of computer programs that will have access to and use data to learn about themselves. Unsupervised learning can be a type of machine learning algorithm used to draw inferences from datasets consisting of an input file without a defined answer. The most popular unsupervised learning approach is cluster analysis, which is used for exporatory data analysis to find hidden patterns or groupings in data.

Clustering is a mechanism that divides a given data set into a homogeneous group based on a specific criterion. Similar objects are kept in a group while different objects are kept in a number of groups. Cluster position and important role in a variety of fields including image processing, mobile communication, computational biology, medicine and economics. The main objective of this paper is to provide a segmentation algorithm that will be used effectively in the image classification method. Another goal is the pattern recognition that we combine an effective clustering algorithm called the K-Mean clustering algorithm.

What is K-Mean?

K-means clustering is one of the only and most common unsupervised machine learning algorithms. The goal of the k-means is simple: group similar data points together and find the underlying patterns. To achieve this goal, k-means searches for a fixed number (K) of clusters in a dataset.

A cluster refers to the set of aggregated data points due to certain similarities

- Technical analysis of exploratory results.
- Implements a non-hierarchical method for grouping objects
- Determines the centroid using the Euclidean method for calculating distances
- Groups of objects defined by a minimum distance.

K-MEANS ALGORITHM

Machine learning with K-means algorithm (Fig 1) is broken down into six steps as follows

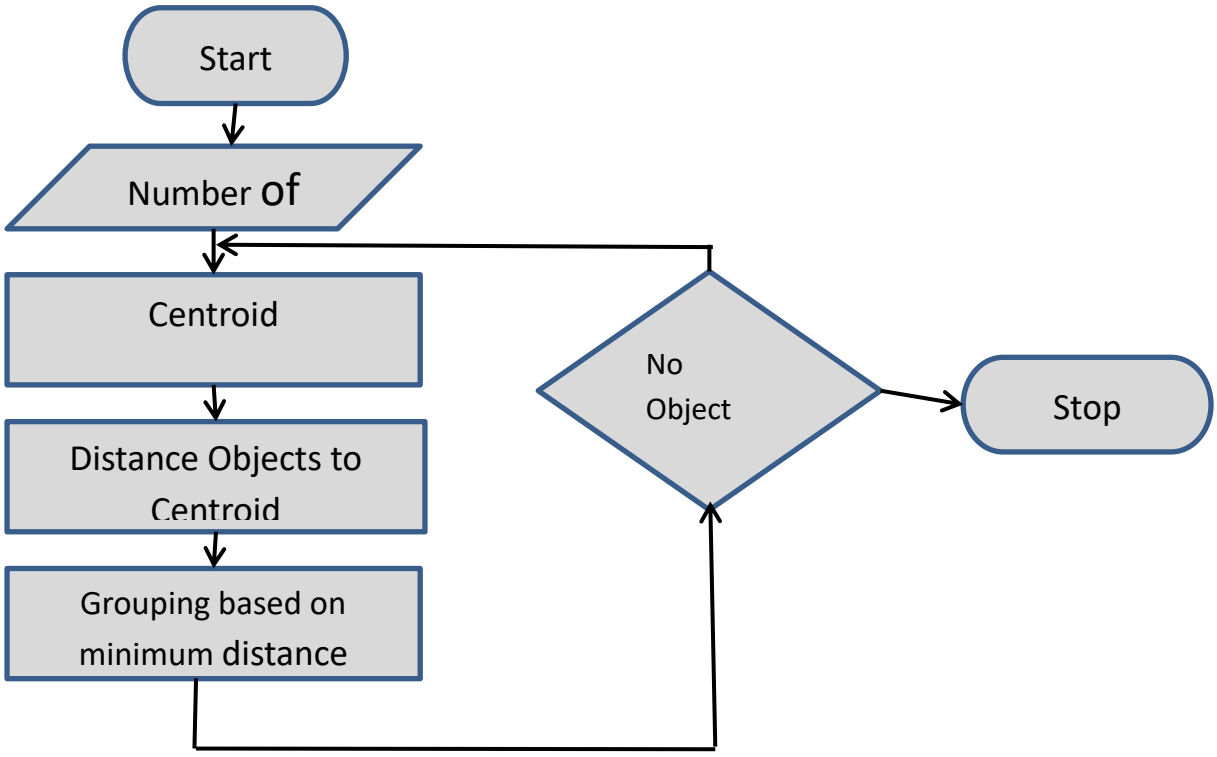

Fig 1 K Means Algorithm

Step 1: assume the two-mean value for the cluster.

Step 2: Calculate the distance using the Euclidean distance formula.

$$Distance[(x,y),(a,b)] = \sqrt{(x-a)^2+(x-b)^2}$$

$$Distance[(x,a)] = \sqrt{(x-a)^2}$$

Step 3: Table the data in respect to the cluster.

Step 4: View the Step5 cluster: recalculate the mean for the new cluster and repeat Step 2 & 4

Step 5: Similar repetitive clusters are formed, then stopped.

Other clustering algorithms with better feature tend to be more expensive. In this case, k-means becomes a great solution for pre-clustering, reducing the space where other clustering algorithms can be applied. K-means is the simplest.

K-Mean Advantages

If variables are massive, then K-Means most of the times computationally faster than hierarchical clustering, if we hold k small. K-Mean produce tighter clusters than hierarchical clustering, particularly if the cluster are globular.

K-Mean Disadvantages

Hard to predict meaning K. It didn't work well with national clusters. Different initial partitions can lead to different final clusters. Doesn't work well with clusters of different sizes and densities (in the original data).

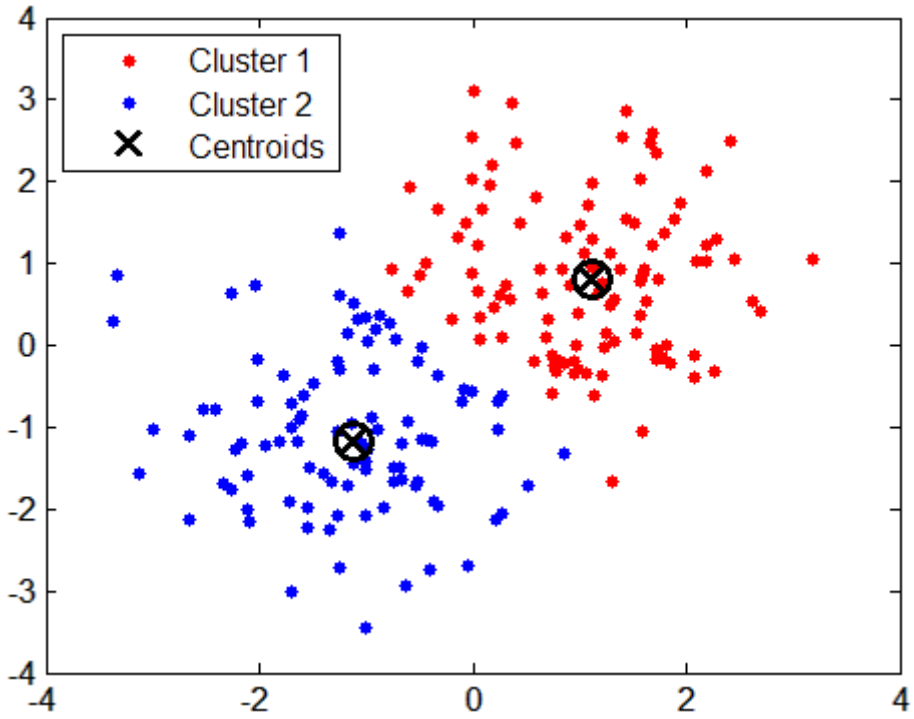

Fig 2 K Means Clustering

IMAGE CLASSIFICATION

The following algorithm outlined the step-by-step approach to the proposed processes of image recognition and segmentation:

(1) Image processing is the very initiative that needs to capture a picture with the aid of a camera.

(2) Input image pre-processing to improve the image quality and remove the unwanted distortion from the image. Clipping the image of the leaf is done to urge the interested image area, then smoothing the image using the smoothing filter is completed. to extend the contrast Image enhancement is additionally done.

(3) Mostly green colored pixels, during this step, are masked. In this, we computed a threshold value that's used for these pixels. Then within the following way mostly green pixels are masked: if pixel intensity of the green component is a smaller amount than the pre-computed threshold value, then zero value is assigned to the red, green and blue components of this pixel.

(4) Detach masked cells within the borders of the contaminated clusters.

(5) Obtain useful segments for the classification of leaf diseases. Segment of components using a genetic algorithm the search functionality of GAs is often used to line unmarked points in N-dimensions into K clusters. We applied an analogous concept to image data in our proposed scheme. We've taken a color image of size m n and each pixel has Red, Green and Blue components. That chromosome has a response, which can be a sequence of K cluster centres. Population is distributed in different rounds and the strongest chromosome survives in each round for subsection. In the fitness computation initiative, the pixel dataset is clustered in accordance with the nearest respective cluster centre, so that each pixel xi of the color image is placed in the respective cluster.

(6) The color co-occurrence approach is used to measure the characteristics using the color co-occurrence technique for the extraction function of the tactics used. It is the approach in which both the look and the color of the image are considered, to get to the unique features that display the image. Over the standard gray-scale representation within the light spectrum, the use of color image features provides an additional aspect of the image characteristics. There are three main mathematical processes in the color co- occurrence system. Second, the conversion of the RGB images of the leaves is done in HIS color space representation. After completion of this process, a color co-occurrence matrix is used for each pixel map, which results in three color co-occurrence matrixes, one for each of H, S, I.

Features called texture features, which include local homogeneity, contrast, cluster shade, strength, and cluster prominence, are computed for the H image as shown in the Est.

(7) Disease classification: during this step of classification, extraction and comparison of the co-occurrence characteristics of the leaves with the corresponding characteristic values are stored in the dataset of the function. Next, the minimum distance criterion and then the SVM classifier are used to complete the classification. The calculation of the progress of the classification shall be completed by the use of the classification benefit and the following. Number of valid classifications=Total number of test images 100

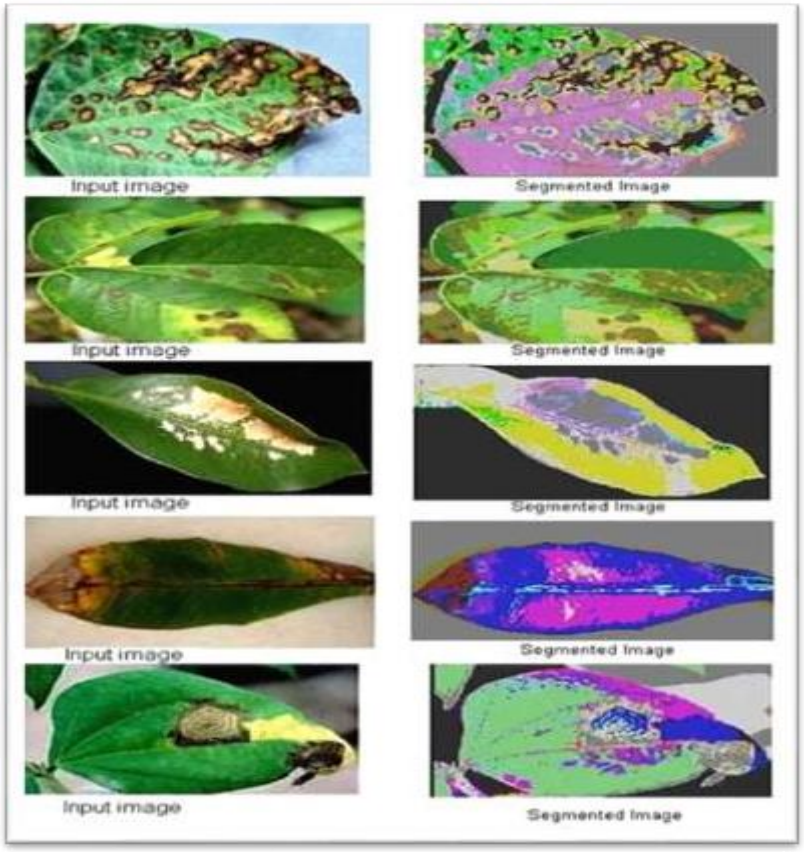

Fig 3 Leaf Detection

RESULT

All experiments are conducted at MATLAB and C#.Net. For input file disease, samples of plant leaves such as rose with bacterial disease, beans leaf with bacterial disease, lemon leaf with sunburn disease, banana leaf with early scorch disease and fungal disease in beans leaf are considered. Fig 4 displays the first images accompanied by the output of the segmented images. Segmented picture is often categorized into a number of plant diseases. Fig 5 displays the input and output image where the input image can be a banana leaf with early scorch disease and the output image shows the disease classification using a feature extraction method.

The classification of diseases of other input plant leaves is shown in the same way. Fig 4.

The co-occurrence features are calculated after mapping the R, G, B components of the input image to the thresholder images. The co-occurrence features for the leaves are extracted and compared with the corresponding feature values that are contained within the function library. Classification is first performed using the K-Mean Clustering Minimum Distance Criterion and has an accuracy of 86.54 per cent. The detection accuracy of the proposed algorithm is increased to 93.63 percent. The second phase classification shall be completed using the SVM classifier and shall show an accuracy of 95.71 percent of its performance. The detection accuracy of SVM with the proposed algorithm is now increased to 95.71 percent. The training and therefore testing collection for each type of leaf in addition to their detection accuracy is shown in Table 1 and from the results, it is often seen that SVM increases the detection accuracy of the classification algorithm compared to other approaches recorded in [3,4,6].

The number of plant disease samples that were categorized into five groups of plant disease using the proposed algorithm is shown in Table 1. It is often seen from the findings that only a few samples of Frog eye leaf spots and bacterial leaf spots have been misclassified. Just two leaves with bacterial leaf spot disease are known as a frog leaf spot and one frog leaf spot.

Fig 4 Input and output image of banana leaf

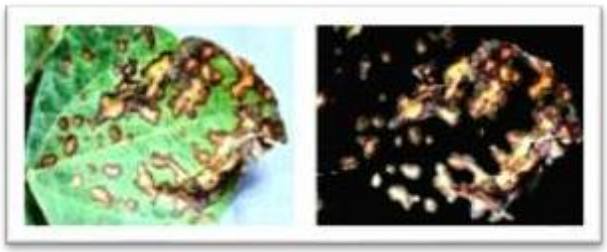

Fig 5 Input and output image of beans leaf

Fig 6 Input and output image of rose leaf

The detection accuracy of SVM with the classification algorithm is now increased to 95.71 percent. The training and therefore testing collection for each type of leaf in addition to their detection accuracy is shown in Table 1.

Table1 Comparison of result

Disease samples	No. of images used for training	No. of images used for testing	Detection accuracy/% MDC with K mean Algorithm	MDC with Algorithm	SVM with algorithm
Banana	15	10	80.00	90.00	90.00
Beans	15	14	92.85	92.85	92.85
Lemon	15	10	90.00	100.00	100
Rose	15	12	83.33	91.66	100
Overall Accuracy			86.54	93.63	95.71

CONCLUSION

This paper provides a survey on the different disease classification strategies used for the detection of plant diseases and an image segmentation algorithm used for automated detection of plant leaf diseases later on. Bananas, bananas, jackfruit, lemon, mango, onions, tomatoes and sapota are a variety of the ten species on which the proposed algorithm is evaluated. As a result, similar diseases for these plants have been reported. Optimal results have been obtained with much less computational effort, which also demonstrates the efficiency of the proposed algorithm in the identification and classification of leaf diseases. The advantage of using this approach is that plant diseases are often detected at an early stage or at an early stage. Artificial Neural Network, Bayes Classifier, Symbolic Logic and Hybrid Algorithms can also be used to increase the recognition rate in the classification process.

Appendix

- First, we're going to identify similar images using the k mean algorithm and we're going to determine the optimum number of clusters using the Elbow process.
- Let's look at the dataset, then first I'm loading and processing the images. First, use the Elbow Form. We're going to determine the optimal number for k.
- The graph reveals that the optimal number for k is 4 using the k=4 parameter.

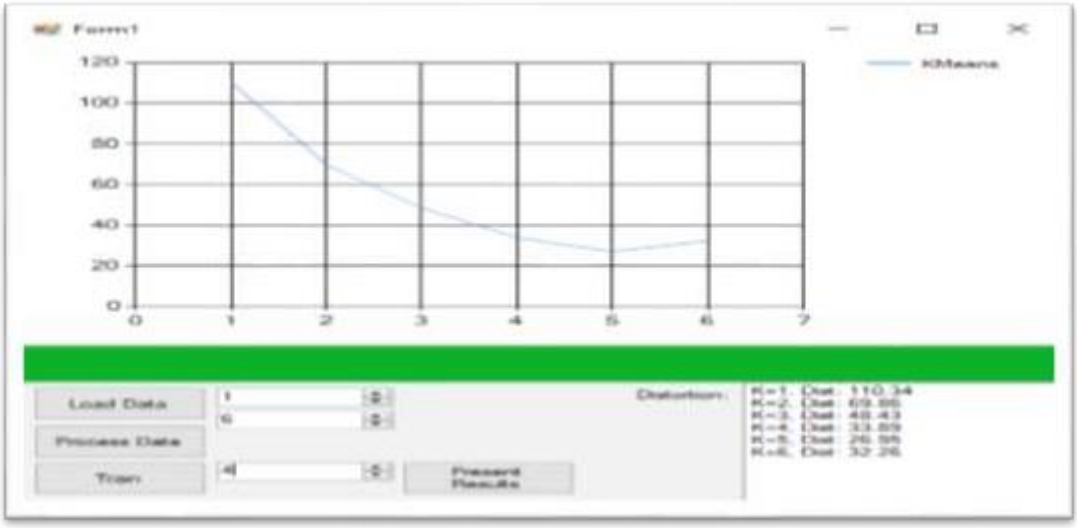

Fig A Elbow Method

Fig B Dataset

Fig C Classify the images using k mean algorithm.

REFERENCES

1. Al-Bashish D, Braik M, Bani Ahmad S. Detection and classification of leaf diseases using k Means-based segmentation and neural-network-based classification.' Inform Technol J 2011; 10:267–75
2. Arivazhagan S, Newlin Shebiah R, Ananthi S, Vishnu Varthini S. Detection of unsanitary plant-leaf regions and classification of plant-leaf diseases by textured surfaces. Agric Eng Int CIGR 2013;15(1):211–7.

3. Beucher S, and Meyer F. The morphological approach to segmentation: the watershed is transformed. In: Dougherty ER, editor. Photo processing mathematical morphology, vol.
4. Twelve. New York: Marcel Dekker, 1993. pp.433–81
5. Bhanu B, Ming J, Lee S. IEEE Trans Syst Man Cybern Dec 1995;25:1543–67. Adaptive image segmentation using a genetic algorithm
6. Peng J, Bhanu B. Adaptive automated segmentation of images and identification of artifacts IEEE Trans Syst Man Cybern Part C 2000;30:427–41.
7. Dhaygude Sanjay B, Kumbhar Nitin P. ' Agricultural plant leaf disease identification using image processing.' Int J Adv Res Electr Electron Instrum Eng. 2013;2(1)
8. Ghaiwat Savita N, Arora Parul.' Identification and classification of plant leaf disease using image processing techniques: a summary.' ISSN (Web).
9. Anand H Kulkarni, Ashwin Patil RK. Applying imaging technique to identify diseases of plants. Int J Mod Eng Res 2012;2(5):3661–4.
10. Amoda Niket, Smita Naikwadi. “Photo processing advances for the identification of plant diseases.

How to cite this article:
R Shalini, C Pradeepa, A Emima, “A Study on Integrated K Mean Algorithm Using Image Classification”, International Journal of Intelligent Computing and Technology (IJICT), Vol.5, Iss.2, pp.63-73, 2022