International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 826-831 www.ijaem.net ISSN: 2395-5252

K-Means Clustering Algorithm for Medical Images

Shaymaa Abdulelah Shaban, Dr.Waleed A. Mahmoud Al-Jawher

Informatics Institute for Postgraduate Studies, Baghdad, Iraq
College of Engineering, Uruk University, Baghdad,Iraq

Submitted: 15-11-2022 Accepted: 26-11-2022

ABSTRACT-
The K-means clustering technique is one of the most popular options for cluster analysis due to its ease of use and quick convergence. A k-value for clustering must be specified in advance, and this parameter has a significant bearing on the final convergence result.

One of the most important tools in data mining is the clustering analysis method, and the clustering results will be affected in direct proportion to the clustering algorithm used. This paper describes the commonly used k-means clustering algorithm for medical images and how each iteration involves analyzing and calculating the distance between each medical image and all cluster centers. The k-means approach has improved clustering speed and accuracy while decreasing computational complexity in experiments.

Keywords- clustering analysis; k-means algorithm; distance; computational complexity

I. INTRODUCTION

Clustering, a form of unsupervised learning, requirements and meets data into groups without an expert present [1]. It's the go-to method for categorizing information into groups with similar characteristics. Unsupervised machine learning is used to find the clusters, representing the hidden patterns. Clustering analysis algorithms and approaches give fundamental tools for processing a variety of uses, including information retrieval, text mining[2], weblog analysis[3], etc. For partitional clustering algorithms to generate qualitative clusters, the number of clusters to create and the location of the initial seed point isof critical importance. If a value of K is known in advance, the K-means algorithm can efficiently organize and analyze massive data [4, 5]. The various automatic approaches for determining the optimal value of K are indicated by literature reviews [5, 6, 7].

II. K-MEANS ALGORITHM

K-means clustering, first developed in 1956 [9] (MacQueen[10,8]), is a simple unsupervised learning approach for resolving the well-known clustering problem. The method assumes a fixed number of clusters (k clusters) to classify a given data set in a straightforward and basic fashion. The key concept is to identify k-cluster centroids. These centers of mass require deft placement, as moving them around can alter the outcome. Therefore, setting them up as far apart as possible is preferable. The next thing to do is to find the centroid closest to each point in the data set. An early grouping is done when there is no waiting point, which means the first step has been completed. Finally, we need to determine k new cluster centroids based on the clusters we just created. Once we have these k new centroid locations, we must be rebinding the original data set points to their nearest new centroid. There is now an infinite cycle. We might see that the positions of the k centroids gradually shift until the loop terminates. So, it's safe to say that centroids are staying put for good.

The final objective of this approach is to reduce the squared error as much as possible.

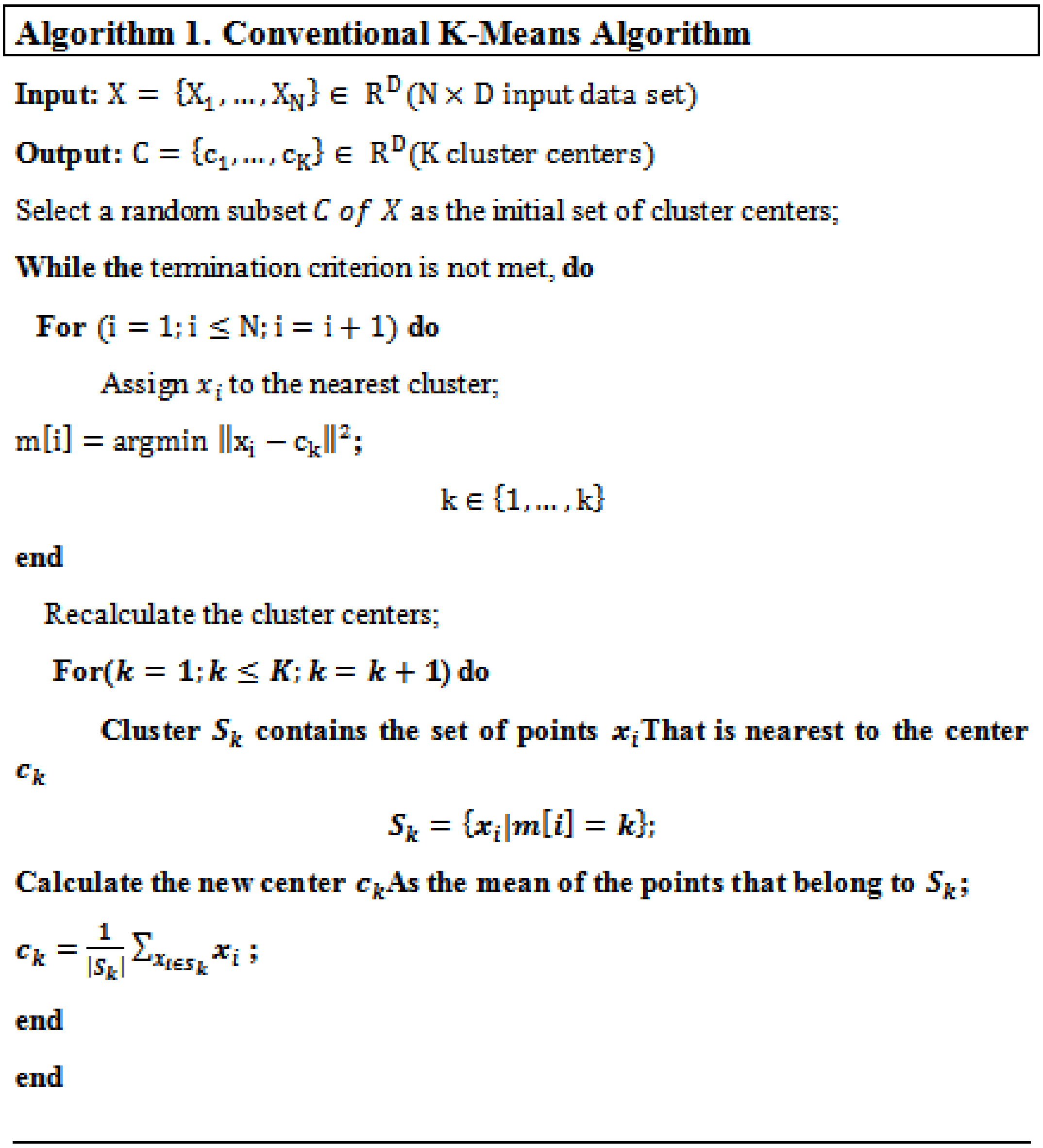
The objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^j - c_j \right\|^2,$$

Where $\left\| x_i^{(j)} - c_i \right\|^2$ is a chosen distance measure between a data point$x_i^{(j)}$ and the cluster centrec_j, is an indicator of the distance of the n data points from their respective clustercenters.

The k-means approach can be shown to converge to a solution, but it is not guaranteed to find the configuration that corresponds to the global goal function's minimum. The algorithm also relies heavily on randomly selecting the starting points for the clusters. Consequently, the k-means technique can be employed repeatedly to reduce the effect of the problem.

The steps that make up the K-Means [11] algorithm are:
1. Choose k random points as the initial centroids.
2. Pick one of the data points and compare it to each centroid. If the data point and the centroid are similar, put it in the same cluster as the centroid.
3. After putting each data point into one of the clusters, recalculate the values of the cluster centers for each k number of clusters.
4. Keep doing steps 2 and 3 until no data point moves (termination requirement met).



Algorithm 1. Conventional K-Means Algorithm

```
Input: $X = \{X_1, \dots, X_N\} \in R^D (N \times D$ input data set)
Output: $C = \{c_1, \dots, c_K\} \in R^D$ (K cluster centers)
Select a random subset $C$ $of$ $X$ as the initial set of cluster centers;
While the termination criterion is not met, do
  For $(i = 1; i \le N; i = i + 1)$ do
        Assign $x_i$ to the nearest cluster;
$m[i] = \text{argmin} \|x_i - c_k\|^2$;
                                   $k \in \{1, \dots, k\}$
end
   Recalculate the cluster centers;
    For$(k = 1; k \le K; k = k + 1)$ do
        Cluster $S_k$ contains the set of points $x_i$That is nearest to the center
$c_k$
                                   $S_k = \{x_i | m[i] = k\}$;
Calculate the new center $c_k$As the mean of the points that belong to $S_k$;
$c_k = \frac{1}{|S_k|} \sum_{x_i \epsilon S_k} x_i$ ;
end
end
```

K-Means Algorithm Flowchart

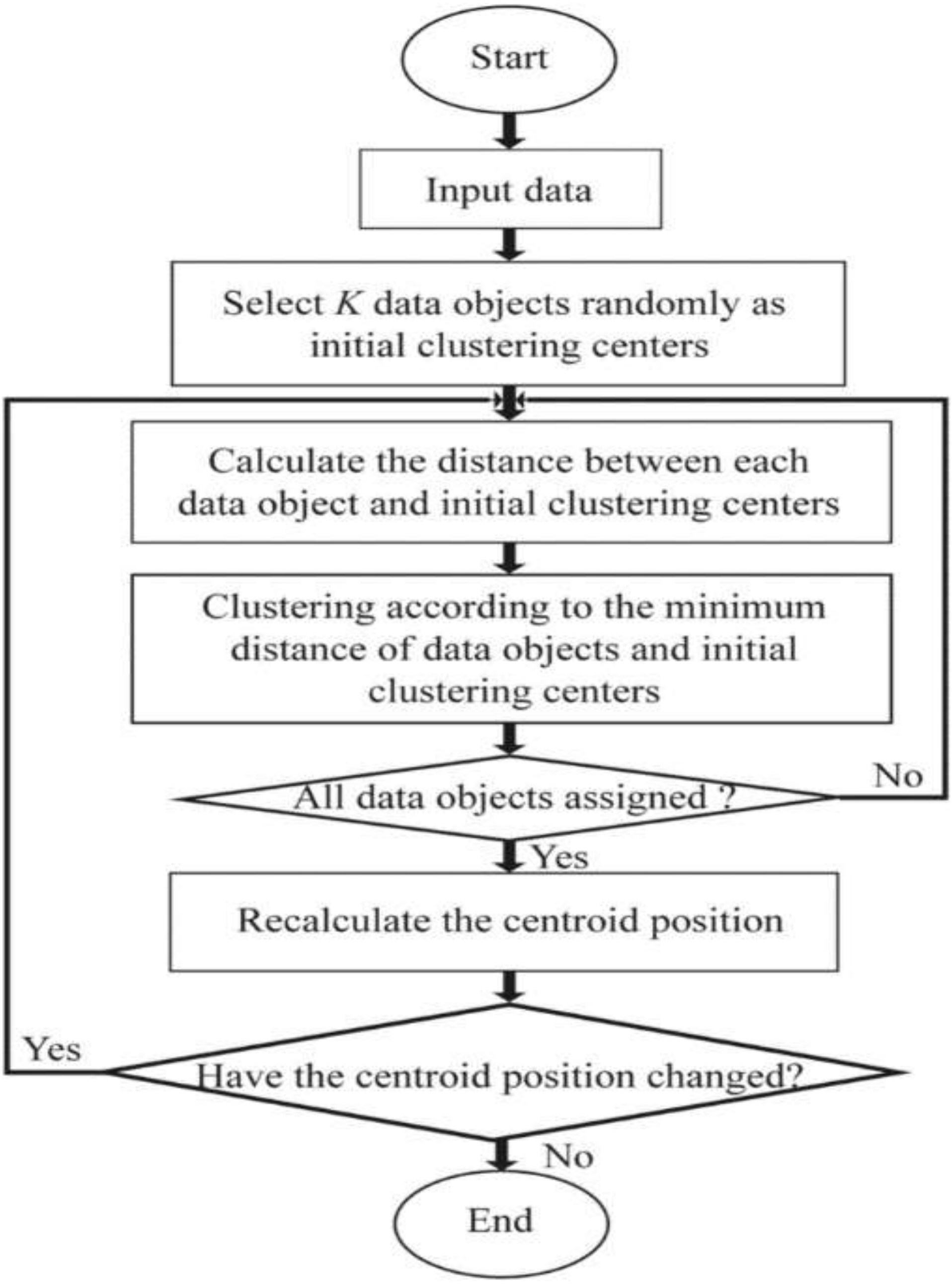

Start
Input data
Select K data objects randomly as initial clustering centers
Calculate the distance between each data object and initial clustering centers
Clustering according to the minimum distance of data objects and initial clustering centers
All data objects assigned ?
No
Yes
Recalculate the centroid position
Yes
Have the centroid position changed?
No
End

Figure (1) k-means Algorithm Flowchart

III. IMAGE CLUSTERING

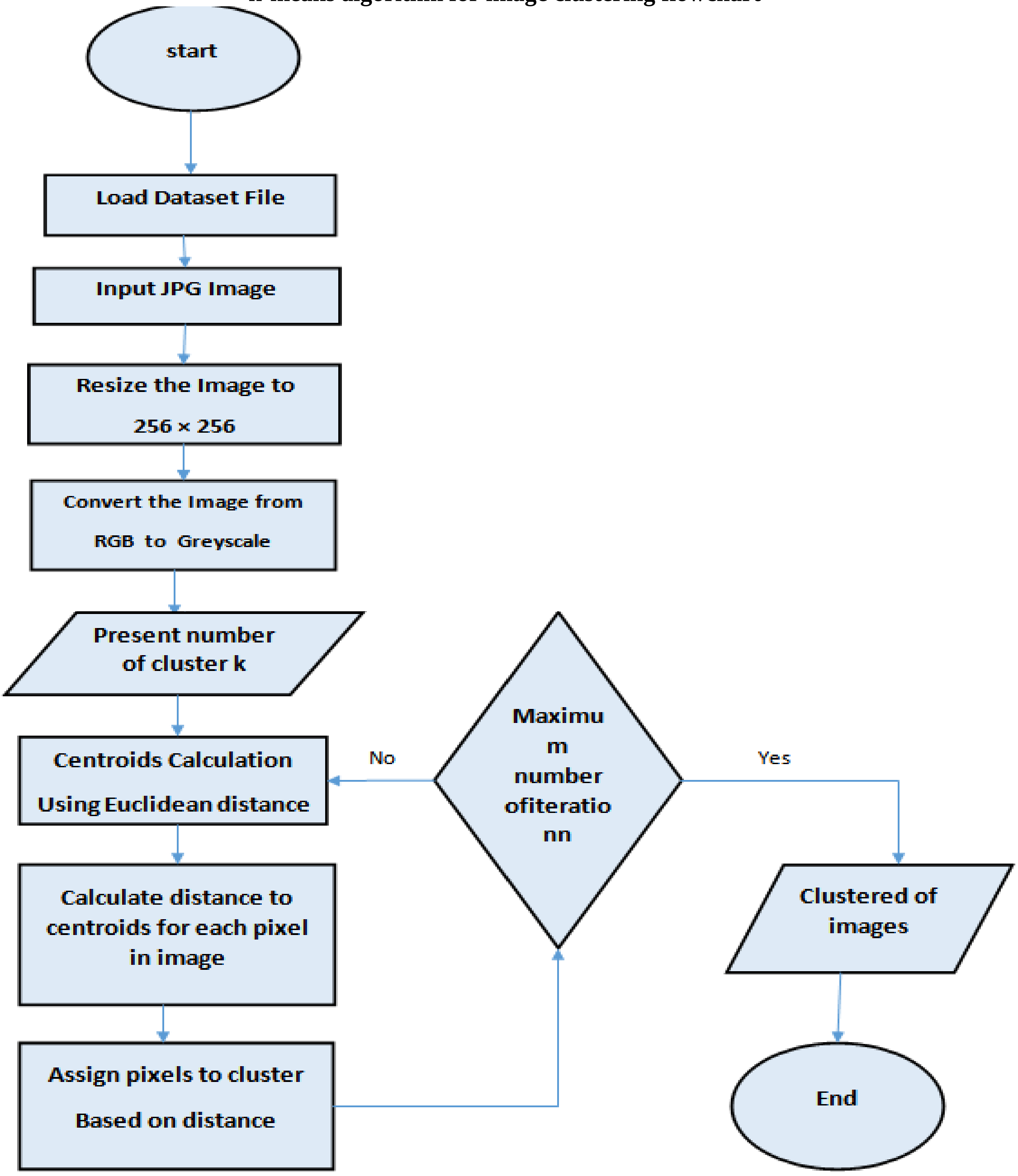
In data mining and image processing, clustering is a significant stage in classifying large amounts of numerical and image data. This method organises the images into a fixed number of groups. Pixels in images are analyzed and classified according to characteristics, including color, texture, and shape.

K-means Algorithm for Medical Image

Pre-processing There is too many problems with the raw images gathered from the scan center and the websites to use for direct processing. Therefore, processing the images before use is required. In magnetic resonance imaging (MRI), labelling, artefact removal, enhancement, and segmentation all benefit from preprocessing [12]. Preprocessing steps like resizing and removing noise improve image quality so that small details can be seen clearly [13].

K-means is a clustering method that can divide an image into a specified number of clusters. A significant similarity between the data and the cluster center is required for K-clustering [14], [15]. And the cluster size (k), where the clusters' centres were chosen randomly from the dataset as input to the algorithms. The next step is to use the Euclidean distance equation to figure out how far apart each pixel is from each cluster's center, and then to use the same distance and scientific repetition to figure out where each cluster's data should be placed until all of the data has been redistributed.

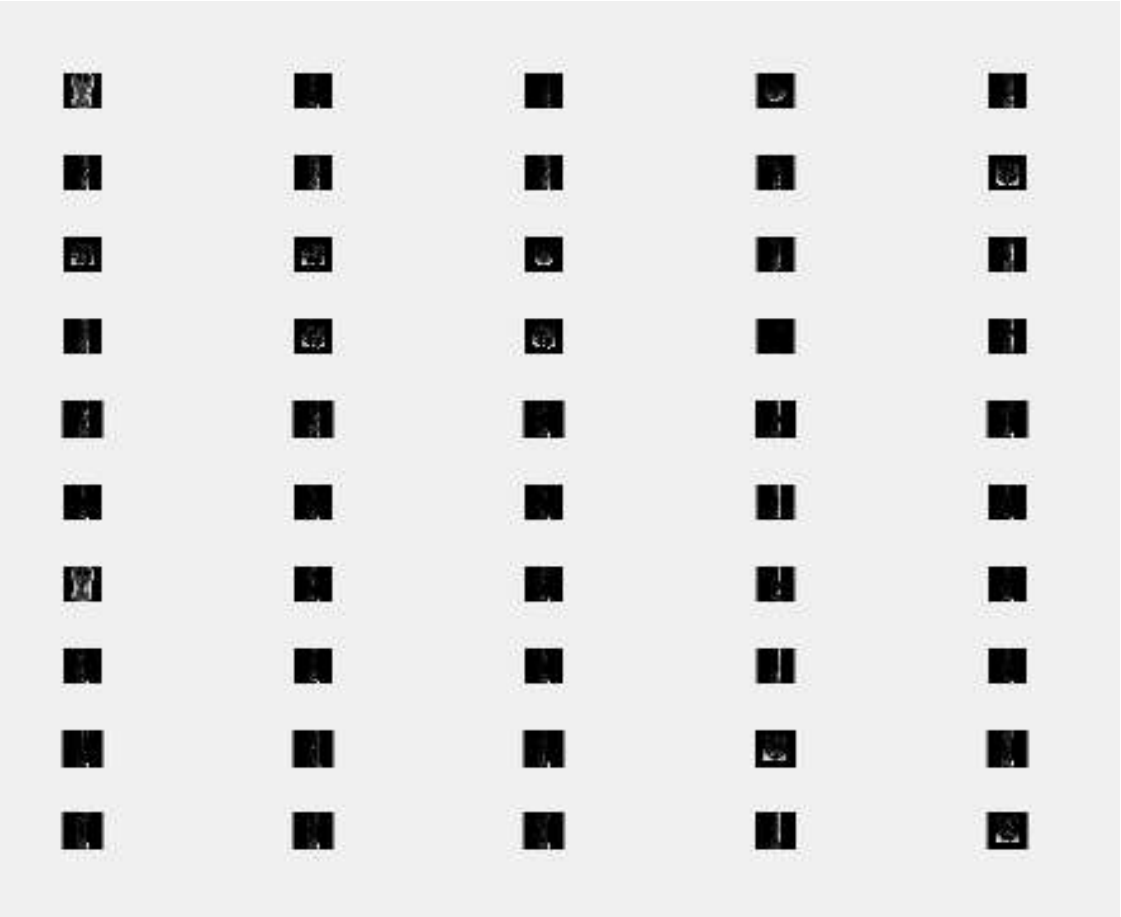
k-means algorithm for image clustering flowchart



Fig(2) k-Means Clustering Algorithm for Medical Images Flowchart

IV. THE IMPLEMENTATION AND RESULT

The MRI image used for implementing the k-means algorithmis real data from Al-Anbar Hospital in Iraq. It consists of 100 images, and the algorithm splits them into five clusters; each cluster contains ten an image depending on the similarities between the images, as shown in figure(3) below.The experiment takes advantage of the MATLAB environment.



Figure(3) clustering of MRI medical image data

V. CONCLUSION

The application of effective data mining methods is currently the primary focus of medical image clustering research. Meanwhile, the primary goal of ongoing research is to enhance the accuracy and computational speed of clustering algorithms while decreasing the need for human interaction. Therefore, this research assesses the famous and traditional approaches to medical image classification. As a result, this article summarises the many medical imaging clustering methods now in use. In addition, the clustering algorithm can aid in making doctors better at detecting and analyzing abnormalities, which leads to more accurate diagnosis and treatment.

REFRENCES

[1]. Jain AK, Dubes RC (1988) Algorithms for clustering data. Prentice-Hall, Inc.

[2]. Bordogna G, Pasi G (2011) Soft clustering for information retrieval applications. Wiley Interdiscip Rev Data Min KnowlDiscov 1(2):138–146.

[3]. Xu S, Qiao X, Zhu L, Zheng H (2010) Deep analysis on mining frequent & maximal reference sequences with generalized suffix tree. J ComputInfSyst 6(7):2187–2197.

[4]. Jain AK (2010) Data clustering: 50 years beyond k-means. Pattern Recognit Lett 31(8):651–666.

[5]. Chen K, Liu L (2005) The ‘‘best k’’ for entropy-based categorical data clustering.

[6]. Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic.J R Stat SocSer B Stat Methodol 63(2):411–423.

[7]. 40. Yadav J, Sharma M (2013) Automatic k-detection algorithm. In: 2013 International Conference on Machine Intelligence and Research Advancement (ICMIRA), IEEE, pp 269–273.

[8]. R.Xu, D.WunschA.Jain, M. Murty, P. Flynn, "Data clustering: A review", ACM Computing Surveys, 31, 1999,pp- 264.

[9]. H. Steinhaus, Sur la division des corp material en parties. Bull. Acad. Polon. Sci., C1, Vol.IV:801-804, 1956.

[10]. J. MacQueen.”Some methods for classification and analysis of multivariate observations”. Proc. of Berkeley Sym. on Math. and Prob., pp-281–297, 1967

[11]. P. Bradley, and U. Fayyad, ―Refining Initial Points for K-Means Clustering,‖ In Proceeding of 15th International Conference on Machine Learning, Jan 1998, pp. 91-99

[12]. Indra K. M., S. N. and S. B. ,”Accurate Breast Contour Detection Algorithms in Digital Mammogram”, International Journal of Computer Applications (0975 – 8887) Volume 25– No.5, July 2011.

[13]. Li, Qiang, and JinghuaiGao. “Contourlet based seismic reflection data non-local noise suppression”. Journal of Applied Geophysics, vol. 95. pp : 16-22,2013.

[14]. U.R. Raval and C. Jani, “Implementing and Improvisation of K-means Clustering Algorithm”, International Journal of Computer Science and Mobile Computing, Vol. 5, No. 5, pp. 191-203, 2016.

[15]. A. Wosiak, A. Zamecznik and K. NiewiadomskaJarosik, “Supervised and Unsupervised Machine Learning for Improved Identification of Intrauterine Growth Restriction Types”, Proceedings of Federated Conference on Computer Science and Information Systems, pp. 323-329, 2016.