# TWO WAY CLUSTERING BASED ON MINIMUM SPANNING TREE AND DBSCAN ALGORITHM FOR IMAGE MINING

Saritha Dasari[1], Dr. A. Rama Mohan Reddy[2], Dr. B. Eswara Reddy[3]
[1]Research Scholar, JNTU, Anantapuramu, India
sarithadasari123@gmail.com
[2]Professor, Department of CSE, S.V. University, Tirupati
[3]Principal, Professor of CSE, JNTU, Kalikiri

## ABSTRACT

Image mining is currently a growing yet active research focus in computer science. Image mining is connected with the development of information mining inside the field of image processing. Image mining handles with the concealed data extraction and additional examples that are not obviously characterized inside the pictures. Image mining incorporates systems like Image preparing, information handling, Robotics and machine learning. Semantic maps are used to visualize the image information which is stored in image databases. But to build the semantic maps we propose one graph optimization technique that is spanning tree techniques. After the development of semantic maps, data mining techniques are used to extract the image information. In this paper we propose a novel algorithm, Two Way Clustering based on Minimum Spanning Tree and DBSCAN (density-based spatial clustering of applications with noise) (TWCMSTDBSCAN) for Image Mining to segment the given image and to detect anomalous outliers (pattern). Minimum Spanning Tree strategies are techniques that produce groups (clusters) by means of graphs. The edges of the graph interface the cases spoke to as nodes. The DBSCAN calculation finds groups (clusters) of self-assertive shapes and is effective for substantial spatial databases. The calculation hunt down groups via looking the area of every item in the database and checks on the off chance that it contains more than the minimum number of objects.

***Keywords:*** *Image mining, semantic map, clustering, spanning tree, DBSCAN.*

## 1. INTRODUCTION

Image mining is a system regularly used to concentrate learning specifically from picture. It uses techniques from computer vision, image processing, image retrieval, data mining, machine learning, database, and artificial intelligence. Ordonez et.al, [1] executed a rule mining ideas for gigantic picture databases. There are two most critical strategies. The principal procedure is to mine from tremendous measure of pictures alone and the second system is to mine from the incorporated accumulations of pictures and related alphanumeric information. Another analyst Megalooikonomou et.al, [2] additionally proposed standard mining procedure to decide relations amongst structures and elements of human mind. Zaiane et.al, [3] proposed a image mining calculation utilizing blob required to be completed the mining of relations inside the connection of pictures.

The principle goal of image mining is to deliver every extensive patterns with no data of the image content, the patterns sorts are distinctive. They could be "classification patterns, description patterns, correlation patterns, temporal patterns and spatial patterns". Image mining handles with all components of immense picture databases which includes indexing strategies, image storages, and image retrieval, all in regards to in a picture (image) mining framework gave by Missaoui et.al, [4]. The foundation of a image mining framework is habitually a complicated procedure since it suggests joining various strategies extending from picture recovery and indexing plans up to information mining and example acknowledgment. Further, it is foreseen that a decent quality picture mining framework furnishes clients with a helpful access into the picture stockpiling territory in the meantime it perceives information designs and creates learning underneath picture representation.

Such framework fundamentally should unite the accompanying capacities: picture stockpiling, picture handling, highlight extraction, picture indexing and recovery and, example and learning disclosure.

Image segmentation is the initial phase in picture (image) mining. Image segmentation is firmly identified with the grouping issue. In Image analysis discovering groups in information is extremely valuable. We can discover pixels with comparable intensities i.e., consequently discovers districts in pictures. We can likewise discover odd articles, which are available in the picture. Division can be seen as segment a given picture into locales or fragments such that pixels having a place with a district are more like each other than pixel having a place with various areas. We likewise require that these locales be associated so districts comprise of touching or neighboring pixels. Countless division strategies are accessible. These strategies depend on one of the accompanying three methodologies (i) clustering (ii) boundary deduction (iii) region growing. Image segmentation has the same relationship to picture characterization. In this paper we utilize two way clustering for portioning (segmenting) pictures. The proposed calculation enhances the execution of various classifiers and decreases the quantity of elements.

One of the best known issues in the field of data mining is grouping (clustering). The issue of clustering is to segment an information set into gatherings (groups) in a manner that the information components inside a group are more like each other than information components in various clusters [5].

Figure 1 demonstrates a general structure model for image mining System. The framework considers a predefined test of pictures as information, whose picture components are removed to speak to briefly the image content. Other than the importance of this mining errand, it is vital to consider invariance issue to some geometric changes and power regarding clamor and different mutilations in planning an element extraction administrator. In the wake of speaking to the picture content, the model depiction of a given picture - the right semantic picture elucidation - is acquired. Mining results are acquired in the wake of coordinating the model portrayal with its integral typical depiction. The typical depiction may be only an element or an arrangement of elements, a verbal portrayal or expression keeping in mind the end goal to and Feature Extraction Mining Interpretation and Evaluation Knowledge Image Database.

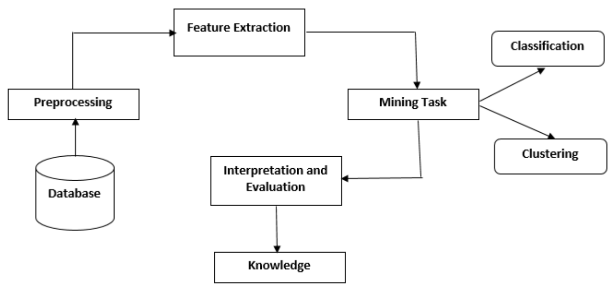


Fig. 1 General structure of Image Mining process

## 2. RELATED WORK

Feature subset selection is a procedure of finding and disposing of however many unessential and repetitive components as much as could be expected under the circumstances. Purpose behind this is 1) insignificant elements don't similar sounding word usage to the prescient precision, and 2) repetitive elements give the vast majority of the data which is as of now present in alternate elements, so it doesn't redound to getting great indicator. Quick calculation dealing with both insignificant and excess components.

Customarily, Feature subset selection is utilized to discover the important elements. A surely understood case is Relief. Alleviation is not reliant on heuristics, it requires just direct time of number given components and preparing cases. For two prescient Relief is exceedingly connected elements yet it is not effective for excess component disposal. Proposed a few upgrades in the calculation which is called as Relief-F. It work with fragmented information set and summing it up to multi-class issues, yet at the same time it can't discover repetitive components. Precision and rate of learning calculation can be influenced.

A Correlation Feature Selection (CFS) assesses subsets of elements on the premise of perceptions that a decent element subset contains includes profoundly associated with the destination yet uncorrelated with each other. Fast Correlation Based Filter (FCBF) distinguishes both pertinent and excess components without pair insightful relationship investigation. Unique in relation to these calculations, FAST calculation works in view of grouping based strategy to choose highlights. Various leveled bunching is a procedure of word determination technique with regards to content characterization. Systems for various leveled bunching fall into two sorts: agglomerative and divisive. Agglomerative various leveled grouping used to expel repetitive components. Quick calculations group the components by utilizing Minimum Spanning Tree strategy.

Liu et.al, [6] proposed a locale level semantic mining approach. As it is simpler for clients to comprehend picture content by district, pictures are divided into a few sections utilizing an enhanced division calculation, each with homogeneous ghastly and textural qualities, and after that a uniform area based representation for every picture is fabricated. Once the probabilistic relationship among picture, area, and concealed semantic is developed, the Expectation Maximization technique can be connected to mine the shrouded semantic.

Wang et.al, [7] tackle the issue of semantic crevice by mining the definitive component designs. Intriguing calculations are produced to mine the conclusive element examples and develop a principle base to naturally perceive semantic ideas in pictures. A precise execution study on extensive picture databases containing numerous semantic ideas demonstrates that the proposed strategy is more powerful than some already proposed strategies.

Zhang et.al, [8] proposed a picture order approach in which the semantic connection of pictures and numerous low-level visual elements are mutually abused. The connection comprises of an arrangement of semantic terms characterizing the classes to be related to unclassified pictures. At first, a multi-target enhancement strategy is utilized to characterize a multi-highlight combination model for each semantic class. At that point, a Bayesian learning technique is connected to determine a setting model speaking to connections among semantic classes. At long last, this connection model is utilized to derive object classes inside pictures. Chosen results from a far reaching test assessment are accounted for to demonstrate the adequacy of the proposed approaches.

Abu et.al, [9] used the Taxonomic Data Working Group Life Sciences Identifier vocabulary to speak to our information and characterized another vocabulary which is particular for commenting on monogenean haptoral bar images (MHBI) to build up the MHBI philosophy and a consolidated MHBI-Fish ontologies. These ontologies are effectively assessed utilizing five criteria which are clarity, cognizance, extendibility, cosmology duty and encoding inclination.

The MST clustering algorithm has been widely used in practice. Xu et.al, [10] use an MST to represent multidimensional gene expression data. They point out that an MST-based clustering algorithm does not assume that data points are grouped around centers or separated by a regular geometric curve. Thus the shape of a cluster boundary has little impact on the performance of the algorithm. They describe three objective functions and the corresponding clustering algorithms for computing a k-partition of the spanning tree for any predefined k>0. The first algorithm simply removes the k −1 longest edges so that the total weight of the k subtrees is minimized. The second objective function is defined to minimize the total distance between the center and each data point in a cluster. The algorithm first removes k −1 edges from the tree, which creates a k-partition. Next, it repeatedly merges a pair of adjacent partitions and finds its optimal 2-clustering solution. They observe that the algorithm quickly converges to a local minimum. The third objective function is defined to minimize the total distance between the "representative" of a cluster and each point in the cluster. The representatives are selected so that the objective function is optimized. This algorithm runs in exponential time in the worst case.

Xu et.al, [11] partition a gray-level image into connected homogeneous regions by constructing an MST from the image. The tree partitioning algorithm minimizes the sum of the variations of the gray-levels of all subtrees, and the gray-levels of two adjacent subtrees are required to be significantly different. Each subtree contains several gray-levels and represents a homogeneous region in the image. Other applications of the MST clustering algorithm in the area of image processing can be found in [12, 13].

Lopresti et.al, [14] suggest an RGB color clustering method by constructing a Euclidean minimum spanning tree. Each distinct color in a given image is considered as a point in the three dimensional RGB color space. Thus each color is a node in the EMST. The weight of an edge is the Euclidean distance between two color nodes in the tree. They compute the average distance of the edges in the EMST once it is built. Subsequently the edges that are "longer" than the average weight by a predetermined amount are removed from the tree, leaving a set of disjoint subtrees. Colors in each subtree are the members of a color cluster. They point out that the EMST based color clustering algorithm may fail when dealing with textures and when there are a large number of colors in an image.

Eldershaw et.al, [15] re-examine the limitations of many 2D clustering algorithms that assume that clusters of a point set are essentially spherical, and provide a broader definition of a cluster based on transitivity: if two points p1 and p2 are close to the same point p0, they are both members of the same cluster. They present an algorithm which constructs a graph using Delaunay triangulation, and remove edges that are longer than a cut-off point. Next, they apply a graph partitioning algorithm to find the isolated connected components in the graph, and each discovered component is treated as a cluster. Unlike Zahn's method in which inconsistency is a locally determined property of an edge, they choose a cut-off point which corresponds to a global minimum.

Sanjay et.al, [16] put forth an image mining technique using wavelet transform. The author proposed an image mining approach using wavelet transform. It uses common pattern identical, pattern identification and data mining models with the intention that a real life scene/image can be associated to a particular category, assisting in different prediction and forecasting mechanisms. It is a three-step procedure i.e. image gathering, learning and classification. Since wavelet transform uses time frequency association, it can be utilized for image mining as a substitute of Fourier transform. Wavelet transform is utilized to decompose an image into dissimilar frequency sub bands and a small frequency sub band is used for Principal Component Analysis (PCA). Classification assists in recognizing the category to which an image relates with. They have constructed a prototype system for identification using DWT + PCA system. The conception of image mining as a consequence can be competently used for weather forecasting so that one can know the natural disasters that may occur in advance.

Image mining approach using clustering and data compression techniques was projected by SabyasachiPattnaik et.al, [17]. Satellite images of clouds play a substantial role in forecasting weather conditions. Frequency of image acquirement ranges from one image per minute to another image per hour based on the climatic environment. These occurrences results in huge collection and creation of image data warehouse. Permanent storage and transmission of images is a demanding task. In their approach, data mining clustering method together with Vector Quantization (VQ) is implemented to cluster and compact static color image. Results are shown to demonstrate the findings both subjectively and visually.

Petra Perner [18] discussed the image mining: subjects, framework, a standard tool and its application to medical-image analysis. A tool and a technique for data mining in picture-archiving systems are provided by this author. It is expected to determine the suitable knowledge for picture examination and identification from the data base of image descriptions. Knowledge-engineering methods are used to acquire a list of attributes for symbolic image descriptions. An expert describes images based on this list and accumulates descriptions in the database. Digital-image processing can be implemented to obtain better imaging of specific image characteristics, or to obtain expert-independent characteristic evaluation. Decision-tree induction is utilized to discover the expert knowledge, provided in the form of image descriptions in the database. This assembled decision tree provides efficient models of decision-making, which can be investigated to maintain image categorization by the expert. A tool for data mining and image processing is developed by this author and its application to image mining is revealed on the task of Hep-2 cell-image categorization. On the other hand, this tool and the technique are standard and can be utilized for other image-mining tasks. They implemented this method in additional medical tasks, for instance, in lung-nodule analysis in X-ray images, lymph-node analysis in MRI and examination of breast MRI.

Lu Kun-Che et.al, [19] projected Decision tree based image processing and image mining technique. Important information can be hidden in images, conversely, few research talks about data mining on them. In their approach, they developed a common framework depending on the decision tree for mining and processing image data. Pixel-wised image characteristics were extracted and changed into a database-like table which permits a variety of data mining algorithms to make explorations on it. Each tuple of the changed table has a feature descriptor produced by a collection of characteristics in conjunction with the target label of a particular pixel. With the label feature, they adopted the decision tree induction in order to comprehend associations among features and the target label from image pixels, and to build up a model for pixel-wised image processing based on a specified training image dataset. Both experimental and theoretical analyses were performed in their study. Their results confirmed that this model can be extremely capable and effectual for image processing and image mining. It is estimated that by using this model, various existing data mining and image processing methods could be worked on together in different ways. Their model can also be used to generate new image processing techniques, enhance existing image processing methods, or act as a powerful image filter.

Jeba Sheela et.al, [20] described the image mining approaches for categorization and segmentation of brain MRI data. Image segmentation plays a vital role in several medical imaging applications by computerizing or assisting the description of anatomical arrangements and additional regions of interest. Automatic recognition of tumors in several medical images is encouraged by the requirement of better accuracy when handling with a human life. Also, the computer assistance is demanded in medical institutions owing to the reality that it possibly will progress the results of humans in such a domain where the false negative cases must be at a very low rate.

It has been confirmed that double reading of medical images possibly will show the way for enhanced tumor detection. But the cost implied in double reading is extremely huge, that's why better software to assist humans in medical institutions is of vast interest at the present time. In their approach they developed a system which uses image mining approaches to categorize the images either as normal or abnormal and then divide the tissues of the anomalous Brain MRI to recognize brain related diseases.

## 3. PROPOSED SYSTEM ARCHITECTURE

The principal phase of the algorithm makes ideal number of group/portions, though the second phase of the calculation further fragments the ideal number of bunches and identify local distribution of closest neighbors. Figure 2 shows the architectural diagram of proposed system.

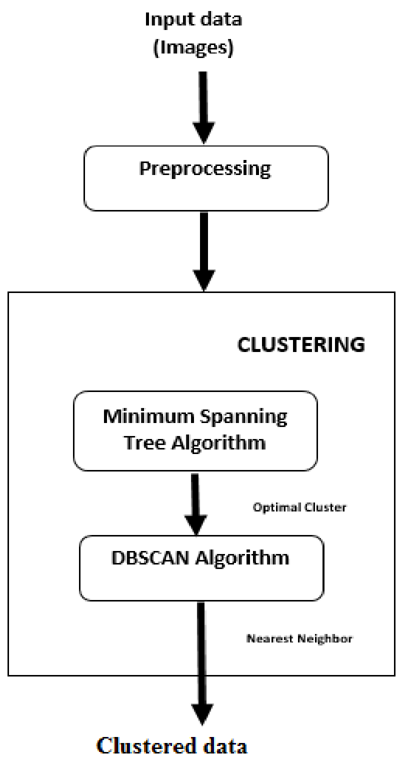


Fig. 2 Proposed system model

### 3.1 Minimum Spanning Tree Algorithm

A spanning tree is "an acyclic sub graph of a graph G, which contains all vertices from G and is also a tree". The minimum spanning tree (MST) of a weighted graph is “the minimum weight spanning tree of that graph” [21].

**Construction of Spanning Tree from Connected Graph**

**Step1:**

Give E a chance to be an edge joining two patient reports in the connected graph G then the weight of an edge (We) is between every pair of document is figured from Jaccard Index as indicated takes after [22].

$$\rho(A,B) = \frac{(|A| \cap |B|)}{(|A| \cup |B|)} - - - - - - - - -(1)$$

Where,

A and B are two picture object.

At that point for every pair discover the distance measure amongst A and B.

W(A,B) = 1- ρ(A,B) ------------------ (2)

This measure can be utilized to calculate distance between categorical attributes. In this way the weight of an edge We between two records A and B.

**Step 2:**

Build a connected weighted graph where the vertices are the pictures and draw the edges between every pair of vertices where weight of every edge is the distance measure between the two relating pictures spoke to as the vertices.

**Step 3: (Identify the number of cycles in the graph)**

For every cycle, expel the most elevated weighted edge required in the cycle from the graph. On the off chance that there exists more than one edge with the same highest value, pick any of them self-assertively (arbitrarily) and expel it from the graph.

### 3.2 DBSCAN Algorithm

The calculation DBSCAN (Density Based Spatial Clustering of Applications with Noise) [23] focusing on low-dimensional spatial information is the significant delegate in this classification. Two info parameters ε and MinPts are utilized to characterize:

1) An ε –neighborhood $N_\varepsilon(x) = \{y \epsilon X \mid d(x,y) \leq \varepsilon\}$ of the point x,

2) A core object is a “point with a neighborhood consisting of more than MinPts points”.

3) A concept of a point y density-reachable from a core object x (a finite sequence of core objects between x and y exists such that each next belongs to an ε neighborhood of its predecessor)

4) A density-connectivity of two points x, y (they should be density-reachable from a common core object).

### 3.3 Proposed Algorithm (TWCMSTDBSCAN)

TWCMSTDBSCAN (k)
Initialize nc ← 1
k gives the desired number of clusters.
Let We be the weight of edge e (as calculated in formula 2)
Let σ be the standard deviation of the edge weights
Let ST = Φ be the set of disjoint sub trees of the Minimum weight spanning tree
Let ε be the neighborhood of the point x.
(phase-1)
Repeat
Construct a spanning tree from the connected graph G comprising of all patient documents. Compute the average weight Wavg of all the edges
Compute the standard deviation σ of the edges
For each e ∈ MST
If We >Wavg + σ
Remove e from MST
nc ← nc + 1
ST = ST ∪ {T} //T is the new disjoint subtree
// If the number of clusters nc is less than k, remove nc − k highest // weight edges so that nc = k
If nc< k
While nc ≠ k
Remove the current highest weight edge from MST
nc ← nc + 1
ST = ST ∪ {T} //T is the new disjoint subtree Return k clusters
// If the number of clusters nc is greater than k
If nc> k
Compute the centroid ci of each Ti∈ ST
ST = U Ti∈ ST {ci}
untilnc = k
Return k clusters
(phase-2)
For each unclassified vertex v ∈V in Ti
Repeat
Extract points which are within the ε distance
If ε >0
Start the clustering process and point is marked as visited
else
this point is labeled as noise
until determine all points in the cluster
Return all clusters
End

## 4. EXPERIMENTAL ANALYSIS OF EXISTING SYSTEM

Fig 3 shows the CS (Cluster separation (CS) is defined as the ratio between minimum and maximum edge of MST) value versus the number of clusters in hierarchical clustering. The CS value < 0.8 when the number of clusters is 5. Thus, the proper number of clusters for the data set is 4. Furthermore, the computational cost of CS is much lighter because the number of sub clusters is small.

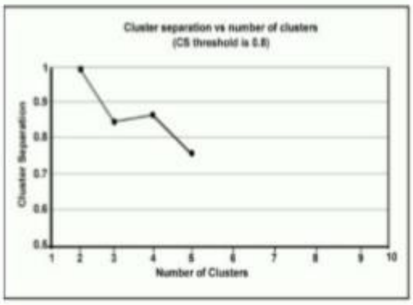


Fig. 3 Number of Clusters vs. Cluster Separation

DBSCAN algorithm have done a good job classifying all the clusters. Nevertheless, it is also superior when comparing the run time between the two algorithms. Graph 1 shows the comparisons, and the resulting differences in time are large. Data taken from paper [24].

To test the efficiency of DBSCAN and CLARANS, we use the SEQUOIA 2000 benchmark data. The SEQUOIA 2000 benchmark database [25] uses real data sets that are representative of Earth Science tasks. The run time comparison of DBSCAN and CLARANS on these databases is shown in table 1.

**Table 1: Run time in seconds**

| No. of Points | 1252 | 2503 | 3910 | 5213 | 6256 | 7820 | 8937 | 10426 | 12512 |
|---|---|---|---|---|---|---|---|---|---|
| DBSCAN | 3.1 | 6.7 | 11.3 | 16.0 | 17.8 | 24.4 | 28.2 | 32.7 | 41.7 |
| CLAR-ANS | 758 | 3026 | 6845 | 11745 | 18029 | 29826 | 39265 | 60540 | 80638 |

The results of our experiments show that the run time of DBSCAN is slightly higher than linear in the number of points. The run time of CLARANS, however, is close to quadratic in the number of points. The results show that DBSCAN outperforms CLARANS by a factor of between 250 and 1900 which grows with increasing size of the database.

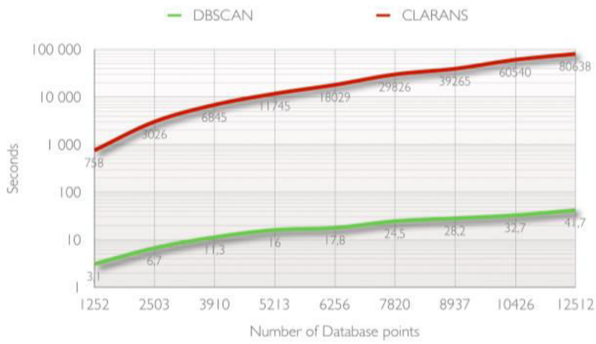


Fig.4 Run time in seconds

## 5. CONCLUSION

The main intent of the image mining is to remove the data loss and obtain the meaningful information which is expected need of human. The proposed two way clustering algorithm works for number of clusters k to be formed and also will consider cases when the number of clusters is not known. Depending on problem specifications, if the parallel solution follows parallel systems basic principles, the future is very promising.

Our proposed algorithm will be implemented in future using Matlab and it will be compared with the existing system.

## REFERENCES

[1] C. Ordonez and E. Omiecinski (1999), “Discovering association rules based on image content,” Proceedings of the IEEE Advances in Digital Libraries Conference (ADL’99).

[2] V. Megalooikonomou, C. Davataikos and E. Herskovits (1999), “Mining lesion-deficit associations in a brain image database,” KDD, San Diego, CA USA.

[3] O. Zaiane and J. Han (1998), “Mining Multimedia Data,” CASCON’98: Meeting of Minds, Toronto, Canada, pp. 83-96.

[4] R. Missaoui and R. M. Palenichka (2005), “Effective image and video mining: an overview of modelbased approaches,” In MDM ’05: Proceedings of the 6th international workshop on Multimedia data mining, pp. 43–52.

[5] S. Guha, R. Rastogi, and K. Shim. (1998), “CURE an efficient clustering algorithm for large databases”. In Proceeding of the 1998 ACM

SIGMOD International Conference on Management of Data, pp. 73-84.

[6] Liu Tingting, Zhang Liangpei, Li Pingxiang, (2012), "Remotely sensed image retrieval based on region-level semantic mining", Eurasip Journal on Image and Video Processing, 2012, Article No. 4.

[7] Wang W, Zhang AD, (2006), "Extracting semantic concepts from images: a decisive feature pattern mining approach", Multimedia Systems, 2006, Vol. 11, No. 4, pp. 352-366.

[8] Zhang Qianni, IzquierdoEbroul, (2013), "Multi-feature Analysis and Semantic Context Learning for Image Classification", ACM Transactions on Multimedia Computing Communications and Applications, Vol. 9, No.2, Article No. 12.

[9] Abu Arpah, Susan Lim Lee Hong, Sidhu Amandeep Singh, (2013), "Semantic representation of monogenean haptoral Bar image annotation", BMC Bioinformatics, 2013, Vol. 14, Article No.48.

[10] Y. Xu, V. Olman, and D. Xu. (2001), "Minimum spanning trees for gene expression data clustering". Genome Informatics, Vol. 12, pp 24–33, 2001.

[11] Y. Xu, E. Uberbacher. (1997), "2D image segmentation using minimum spanning trees." Image and Vision Computing, Vol. 15, No. 1, pp 47–57, 1997.

[12] R. Gonzalez, P. Wintz. (1987), "Digital Image Processing (2nd Edition)". Addison-Wesley, Reading, MA, USA, 1987.

[13] Y. Xu, V. Olman, E. Uberbacher. (1998), "A segmentation algorithm for noisy images: design and evaluation." Pattern Recognition Letters, 19:1213–1224, 1998.

[14] D. Lopresti, J. Zhou. (2000), "Locating and recognizing text in www images." Information Retrieval, Vol. 2, pp 177–206, 2000.

[15] C. Eldershaw, M. Hegland. (1997), "Cluster analysis using triangulation". In B. Noye, M. Teubner, and A. Gill, editors, Computational Techniques and Applications: CTAC97, pages 201–208. World Scientific, 1997.

[16] Sanjay T. Gandhe, K. T. Talele and Avinash G. Keskar (2007), "Image Mining Using Wavelet Transform", Knowledge-Based Intelligent Information and Engineering Systems, Springer link book chapter, pp. 797-803.

[17] SabyasachiPattnaik, Pranab Kumar Das Gupta and ManojranjanNayak (2008), "Mining images using clustering and data compressing techniques", International Journal of Information and Communication Technology, vol. 1, no. 2, pp. 131-147.

[18] Petra Perner, (2002), "Image mining: issues, framework, a generic tool and its application to medical image diagnosis", Engineering Applications of Artificial Intelligence, vol. 15, no. 2, pp. 205-216.

[19] Lu Kun-Che and Yang Don-Lin, (2009), "Image Processing and Image Mining using Decision Trees", Journal of information science and engineering, vol. 25, no. 4, pp. 989-1003.

[20] L. Jaba Sheela, V. Shanthi, (2007), "Image Mining Techniques for Classification and Segmentation of Brain MRI Data," Journal of Theoretical and Applied Information Technology, vol. 3, no. 4, pp. 115-121.

[21] Asano,T., Bhattacharya,B., Keil,M., and Yao,F (1988), "Clustering algorithms based on minimum and maximum spanning trees." In Proceedings of the 4th Annual Symposium on Computational Geometry, pp.252–257.

[22] Satu Elisa Schaeffer (2007). "Survey on Graph Clustering," Elsevier Computer Science Review, pp.27-64, doi:10.1016/j.cosrev.2007.05.001.

[23] Ester, M., Kriegel, H-P., and Xu, X. (1995), "A database interface for clustering in large spatial databases." In Proceedings of the 1st ACM SIGKDD, pp. 94-99.

[24] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, (1996), "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Published in Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96).

[25] Stonebraker M., Frew J., Gardels K., and Meredith J. (1993),"The SEQUOIA 2000 Storage Benchmark," Proc. ACM SIGMOD Int. Conf. on Management of Data, Washington, DC, 1993, pp. 2-11.