ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

Empirical Study of Image Classification Techniques to Classify the Image using SVM: A Review

Anshu Dhabhai, Yogesh Kumar Gupta

M.Tech Student, Department of Computer Science, Banasthali University, Jaipur, Rajasthan, India

Assistant Professor, Department of Computer Science, Banasthali University, Jaipur, Rajasthan, India

ABSTRACT: Medical data are increasing continuously so we need to classify these data and classification is very important task which can help in finding the hidden information from the images and specialist to detect the affected area and disease early. In this paper image classification is our main objective and we try to understand that which type of data we need to classify and what are the classification techniques which we can apply to classify the medical images. Here we review on some classification techniques and we are mainly focus on the SVM (Support Vector Machine) technique. It is widely used today because it handles the multi dimensional data and gives the best and accurate result.

KEYWORDS: Medical Imaging, Data Mining, Neural Network, SVM.

I. INTRODUCTION

The amount of images are increasing, we need to collect them every day so it create problem for us to extract the meaningful hidden information from vast collection datasets of raw images with no human intervention. So we need software's to analyze those images with reliability and high accuracy which helps to reduce the workload [2] and the goal is not only to get result with high accuracy but our goal is to detect the affected area by the disease and help to specialist to learn about the progress of a disease in early [4] .

Classification of the images is very important task to retrieve the knowledge which helps medical specialists in process of decision making. So different kinds of classification methods used for medical image classification the first one is texture classification technique which is used foe image processing to identify the various regions of image on the basis of texture properties. Second one is neural network classification techniques play an important role with including supervised and unsupervised technique. The last one is data mining classification technique, mining means to extract the hidden, useful knowledge from the vast amount of data. It works with statistical, machine learning, virtualization and some techniques of manipulation to extract the knowledge [3].

Here we introduce with some classification techniques which are used in medical images and those techniques are Association rule mining, k-nearest neighbor, texture classification, support vector machine, naïve Bayesian classifier, and multilayer neural network.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

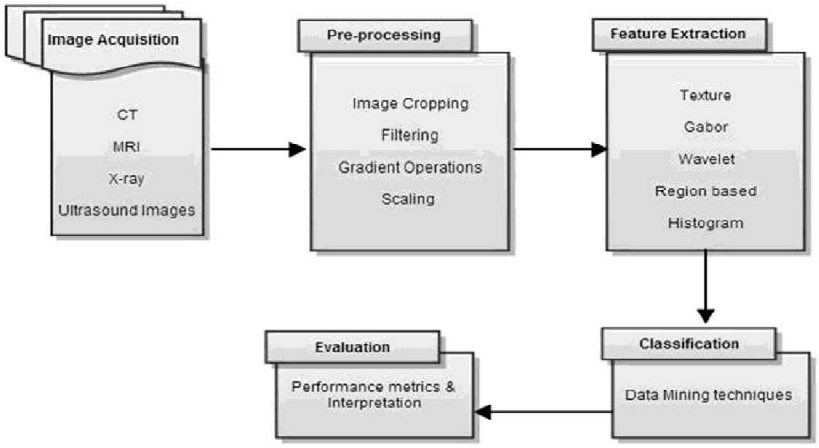

Fig 1. Classification process of medical images [4]

II. LITERATURE SURVEY

In [1] author Introduced with the concepts of medical imaging and the need to classify the medical images. Then discuss the general process of classification, data collection, image pre-processing steps and feature extraction phase. Further they used neural network techniques and the association rule classification technique for classification. At the end they perform the experiment on some data and get result. In [2] author discussed here the complete classification process including "image acquisition, image enhancement, feature extraction, and classification". It uses the apriory algorithm of association rule mining for classification of medical images. In this paper the goal of author is to increase the accuracy of diagnostic and optimize the decision time. In [4] author discuss why we need classification, how can medical images classified using related technologies and methods such as texture classification, k-nearest neighbor, neural network, SVM and describe the proposed framework which consist the six phases "Data acquisition, Data pre-processing, Data partition, Soft set classification, Data analysis and performance calculation" to achieve accuracy and speed of computations. In [5] author introduce with data mining techniques used for classification of medical images. They proposed a new approach for automated classification. Further they proposed discritization method to increase the classifiers accuracy, this method experiments on three classifiers that are K-nearest neighbor, Naïve bayes and SVM. In [6] author explore the various classification techniques that are new and used for medical imaging data such as fuzzy logic, multi layer neural network, polynomial neural network, radial basis function neural network, neuro fuzzy technique, adaptive neuro fuzzy inference system, neuro genetic techniques. Further they explore the phases of medical image data classification. In [7] author review on different kinds of classification techniques for medical image datasets such as Classification and Regression Trees (CART), The Cat Swarm Optimization Algorithm (CS.O), Chi-Squared Automated Interaction Detection, QUEST, Discriminate Analysis, Logistic Regression, Naïve Bayes, Texture Classification, neural network classification, Support Vector Machine (SVM). They proposed a comparison table of different techniques regarding the Accuracy. They also proposed new method for cancer detection using SVM and Neural Network classifier, also used the old technique Nearest Neighbor classifier. In [8] author describes the feature reduction and classification techniques for MRI images. And the good thing is that they mention the various techniques for classification and feature reduction in a tabular form with their name and description.

III. CLASSIFICATION TECHNIQUES

A. *ASSOCIATION RULE MINING USING APRIORI ALGORITHM:*

This approach is good for analyzing the huge set of images. Here we find the particular characteristics of a feature and on the basis of this feature calculate the number of occurrence of that particular feature on the image. Association rules are developed to examine the data to find the frequent patterns. It is use "if/Then" declaration to discover the consociation among the data which are not related and the relational db. For example if user have dozens of maggie then the 80% chances that he is shop the maggie's. This rules works in two parts named as ancedent and consequent.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

Within the data, an item is found is an ancedent and consequent is the item that discover in composition through the ancedent. It uses the two terms support and confidence to recognize the consociation. Support indicates that how many items are frequently appeared in db and confidence gives the indications that how many times "if/Then" statement is true.

Generally it has two phase procedure where in first phase frequent items are finding and another phase is to find association rule between the extract features through the image database. The aim is to find which feature is belongs to which image. After finding all the feature frequent sets according to their category on the basis of apriory algorithm of association rule discovery. When all the features are find they will be merged and saved in transactional db. After that t o find the association rule within db constrained implement the apriory algorithm [1, 2].

B. ***SUPPORT VECTOR MACHINE (SVM):***

SVM is a data manipulation and calculation understanding principle to assess data and to define various type of relationship among the data. Basically the design purpose of SVM is to differentiate the two and more classes. The training methodology used in SVM is to access best linear hyper plane to minimize imminent error due to unknown reasons. SVM is better when it handle multi dimensional but non discrete features. Limitation of SVM, first one is longer response time and inability to deal with isolated attributes, and second one is sample used in it needed larger size to get best outcomes and accuracy in prediction [4, 5, 7 and 10].

C. ***TEXTURE CLASSIFICATION:***

The main objective in texture classification is to allocate given image to one of set of known texture classes. It is to be noted that segmentation of an image is depended on intensity of localize spatial distribution. The real purpose of classification is to detect defects and errors. It is categorized in four parts that are statistical, geometrical, model based and the signal processing. In these days automated pattern recognition and image analysis can be easily done by using modern tools they helps in clinical decision-making and eliminates the variability.

Use of Wavelet in texture classification and analysis, it is designed in such a way that it cannot evaluate distributive frequency on different scales. So it is a multi scale technique for analysis. Whereas Fourier transform can detect the distributive frequency in signal intensity but it is unable to delineate temporal modification in presume and content of frequency [7, 3].

D. ***NAÏVE BAYESIAN (NB) CLASSIFIER:***

Naive bayes classifier is also the best classification methods which work with the simple concept based on conditional probability bayes rule. It assumes that all datasets attributes are independent from each other and have many benefits such as simple, efficient in terms of computation, require less training data and needs all parameter in decision making. Its performance improves with the help of discritization; discretization is a procedure of transform the continuous parameter values into distinct interval gap which increase accuracy, speed and easier understanding [5].

E. ***MULTI LAYER NEURAL NETWORK (MLNN):***

MLNN resolves the complications in classification with the set of for non linear data through invisible layers, where neurons are straightly not related with output. To increase the division competency of network extra invisible layers are explained the arithmetically as supplementary hyper-planes. The BPN (Back propagation network) is the famous MLNN with gradient. BPANN (Back propagation artificial neural network) is the technique of neural network capable to improve the nonlinear values depends on gradient. To improve the network we have a popular algorithm name as EBPA (Error back propagation algorithm) and it has two phases: first one is Feed Forward where the final output is evaluated, and second one is Feed Backward where the evaluated error is broadcast back again to manage weights [6].

F. ***K- NEAREST NEIGHBOR (KNN):***

It is a method of supervised learning whose aim is to accept the k instances, which are nearest with the given instance. Euclidian distance used for evaluating the fixed distance of the k instances. It is also a non parametric method.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

IV. CLASSIFICATION PROCESS MODEL

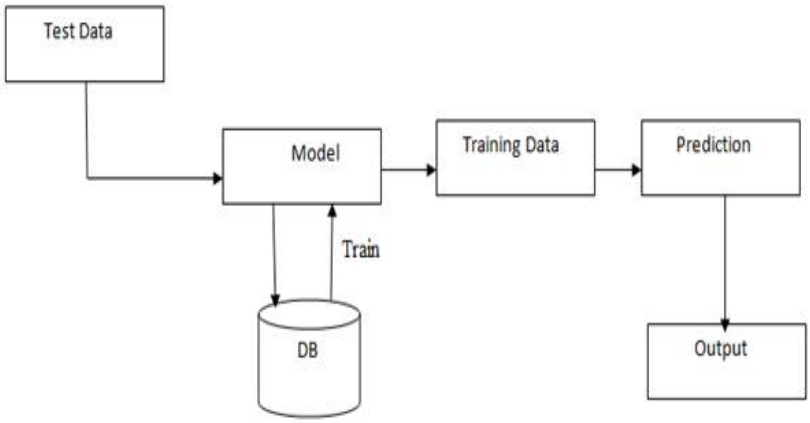

Fig 2. Classification Process Model

In the process of classification we have the Training data that stored in the database so we test the input data called as test data which is match with the data stored in database for the classification process then we get the train data. Then this training data is evaluated and predicted for the final outcome.

V. RESULTS & DISCUSSION

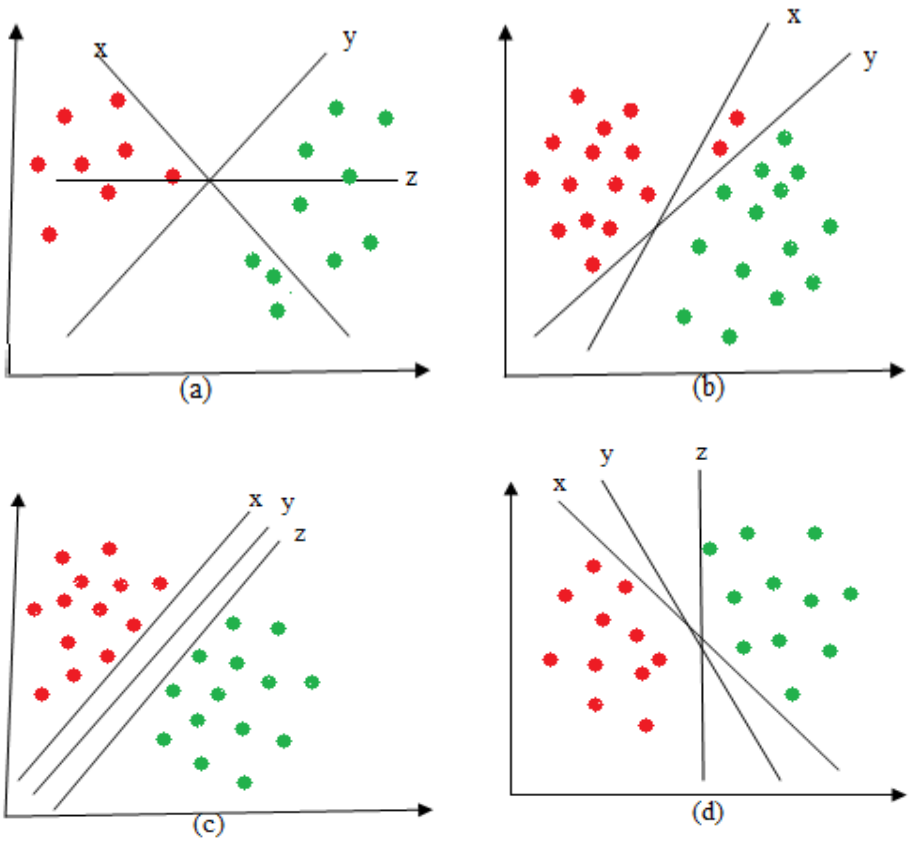

Fig 3. (a), (b), (c) and (d) linearly separable data by SVM

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

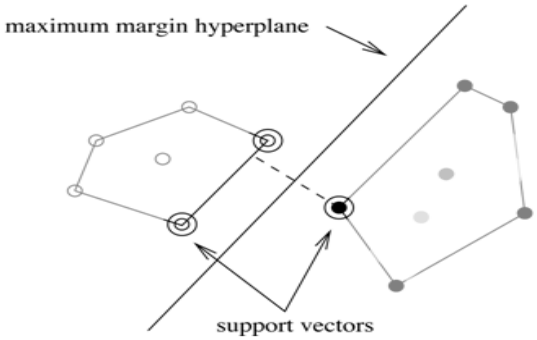

Fig 4. Maximum Margin and Support Vector [17]

When data separate in linear manner then we can make the infinite lines between them but the problem is that we have to select the best optimal line which has the less error in classification. In given figure (d) we can see that the line z and x are highly sensitive so we take y line which is less sensitive. When data is not linearly separable then we transform the primary data into high dimension data through the non-linear mapping.

The goal of SVM is to identify the best hyper plane which classifies the 2 classes. Hyper plane can be positive and negative it is calculated by given formula-

(p.q) + r = 1 (when positive)
(p.q) + r = -1 (when negative)
(p.q) + r = 0 (when hyper plane)

We can find the values pa and r to solve these equations from the linear algebra.

Then in the next step we calculate the margin. Margin is the gap or space between breakable lines to the nearest point of both classes. We select the highest margin line from given formula-

Margin =2/√ (k, k)

After search we find that all the data sets are not classified correctly and sometimes they are not simple, the points are mixed. To handle this problem of classification error we use Support Vector Machine.

$$\varphi(\omega, \varepsilon) = \frac{1}{2}(\omega \,.\, \omega) + C\left(\sum_{i==1}^{l} \varepsilon i\right)$$

$$\text{s.t yi}(< w.xi > -b) \geq 1 - \varepsilon\text{i}$$

$$\text{and i} = 1,2,3 \ldots \ldots l$$

Formula 1. Margin Increasing and Decreasing [17]

If value of C increases then margin becomes maximum and if C is decreases then margin is minimum.SVM is very effective when dimensions are greater than existing samples. SVM use a subset named as Support Vector which classify the various objects. So it is memory efficient, versatile and gives the correct result. But it is poor in probability estimate.

VI. CONCLUSION

Medical image data are increasing day by day so we need to maintain the records of the patient's diagnosis and clinical data record are very difficult and important task. The solution is to classify the images with different techniques which helps to improve accuracy, evaluation speed reduce the manual work, increase the efficiency and quality. In this article we review on some papers and discuss some classification techniques. Here our main focus on SVM technique that are using today to classify the medical images and gives the better and efficient results than the other techniques.

REFERENCES

1. Antonie, M. L., Zaiane, O. R., & Coman, A. (2001). "Application of Data Mining Techniques for Medical Image Classification." MDM/KDD, 2001, 94-101.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

2. GHITA, S. "Using the Apriori Algorithm for Medical Image Classification."
3. Smitha, P., Shaji, L., & Mini, M. G. (2011). "A review of medical image classification techniques." In International conference on VLSI, Communication & Intrumrnataiom (pp.34-38).
4. Lashari, S. A., & Ibrahim, R. (2013). "A Framework for Medical Images Classification Using Soft Set." Procedia Technology, 11, (2013). 548-556.
5. http://www.isaet.org/images/extraimages/D313037.pdf
6. Hota, H. S., Shukla, S. P., & Gulhare, K. (2013). "Review of intelligent techniques applied for classification and preprocessing of medical image data." IJCSI International Journal of Computer Science Issues, 1, 267-272.
7. Chatap, N. J., & Shrivastava, A. K. (2014). "A Survey on Various Classification Techniques for Medical Image Data." International Journal of Computer Applications, 97(15).
8. Kumar, R. S., & Karnan, M. (2014)." Review of MRI image classification techniques." International Journal of Research Studies in Computer Science and Engineering, 1(1), 21-8.
9. Zhou, X., Han, M., Song, Y., & Li, Q. (2013). "Fast Filtering Techniques in Medical Image Classification and Retrieval." In CLEF (Working Notes).
10. D. Jegelevicius, A. Krisciukaitis, A. Lukosevicius, V. Marozas, A. Paunksnis, V. Barzdziukas, et al. (2009). "Network Based Clinical Decision Support System" Proceedings of the 9th International Conference on Info~ation Technology and Applications in Biomedicine, ITAB 2009, Larnaca, Cyprus.
11. Suguna, N. & Thanushkodi, K. (2010). "An Improved k-Nearest Neighbour Classification Using Genetic Algorithm." IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 2, ISSN (Online): 1694-0784.
12. K. M. Al Aidaroos, A. A. Bakar and Z. Othaman. (2012). "Medical data classification with Naïve Bayes approach.", Information Technology Journal, vol. 11, no. 9, pp. 1166 – 1174.
13. R. Agrawal, T. Imielinski, and A. Swami. (1993). "Mining association rules between sets of items in large databases." In Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data, pages 207–216, Washington, D.C., May 1993.
14. Ko, B. C., Kim, S. H., & Nam, J. Y. (2011). "X-ray image classification using random forests with local wavelet-based CS-local binary patterns." Journal of digital imaging, 24(6), 1141-1151.
15. Foody, G. M., & Mathur, A. (2004). "A relative evaluation of multiclass image classification by support vector machines." IEEE Transactions on geoscience and remote sensing, 42(6), 1335-1343.
16. Van den Elsen, P. A., Pol, E. J., & Viergever, M. A. (1993). "Medical image matching-a review with classification." IEEE Engineering in Medicine and Biology Magazine, 12(1), 26-39.
17. Shah, Parin M. (2012). "Face Detection from Images Using Support Vector Machine" (2012).Master's Projects. Paper 321.