IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

A Survey Paper on Data mining Techniques and Challenges in Distributed DICOM

P. Haripriya[1], R. Porkodi[2]

Research Scholar, Department of Computer science, Bharathiar University, Coimbatore, India [1]

Assistant Professor, Department of Computer science, Bharathiar University, Coimbatore, India [2]

Abstract: DICOM images are complex objects, due to the nature of storing clinical data and patient images in a single file. The data sector stores roughly around 2000 tags which comprises of clinical, technical, etc. The Image sector stores multiple images as slices for a particular study. The data and images bind into a single file. It's essential to bind data and image to identify patient records. This makes DICOM images a complex object to store, retrieve and process. Thus we need multiple advanced technologies to solve complex problems in DICOM. By applying multiple Data mining Techniques, information retrieval complexity in DICOM meta data is simplified. Distributed storage is essential for quality data mining. This paper investigates mainly on the data mining techniques used in DICOM (Medical Imaging) which are stored in distributed storage. Data mining on DICOM enables quick retrieval.

Keywords: DICOM, Data mining, Distributed storage, Clustering.

I. INTRODUCTION

DICOM (Digital Imaging and Communications in Medicine) is the ISO 12052 international standard for medical images and related information, evolved around 1980s. [1] ACR/NEMA 300 is the first standard released in 1985 by American College of Radiology (ACR) and National Electrical Manufacturers Association (NEMA). DICOM is introduced in 1993 as the third version of the standard. DICOM gained the international acceptance as a standard in medical imaging. The DICOM standard provides an open source platform with vendor independent for communication of medical imaging and related clinical data. DICOM is a hierarchical object which contains the metrics, data values, image information and image slices. DICOM is a client server model which consists of standard definitions for Data structures, requirements for conforming devices and applications, network services, communication protocols, formats for media exchange. DICOM standard is well established in radiology (CT, MRI, Ultrasound, X-Ray etc). There are almost billions of DICOM images used for clinical care. DICOM has revolutionized the practice of radiology, allowing the replacement of films with digital media. Internet enables distributed storage via cloud computing. Both doctors and patients benefit from these advance DICOM model.

Storing the DICOM images in a distributed environment securely and applying data mining techniques on DICOM are essential for treatment effectiveness, basic healthcare management, best practices, fraud / detection and CAD. Data mining plays critical role in image retrieval and multi level predictions, Data Intelligence. The main objective of this paper is to investigate data mining techniques applied in DICOM in distributed storage. The DICOM Dictionary makes DICOM a structured object [34].

A. DICOM Architecture

The figure 1 shows that the DICOM file stored with the extension ".dcm". Multiple image slices are stored in DICOM. Thus the images in the DICOM shall be viewed as series of image slices. DICOM is not only used to store images. DICOM stores reports, ECG Signals and Audio. DICOM viewer software is used to visualize the entire sequence of images as a movie with corresponding tags. The DICOM file consists of more than 2000 tags.

Fig. 1 DICOM Image (Thox.dcm)

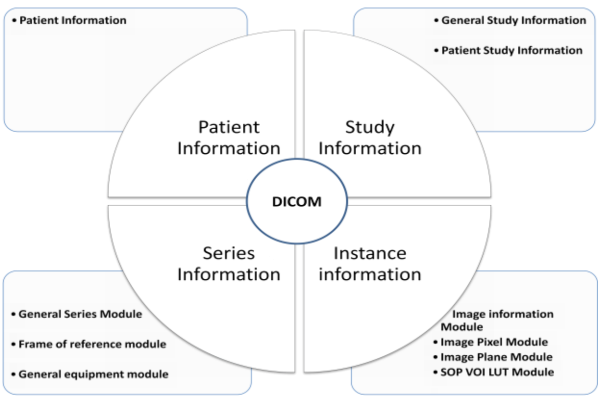

Fig. 2 DICOM Architecture

The figure 2 shows that the DICOM File contain four types of information module such as Patient information,

 DOI 10.17148/IJARCCE.2016.53179

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

Study information, Series information, Instance information. The Patient Information module is used to store the patient personal details such as Patient id, Patient Name, Patient Address, Age, Sex, etc. The Study Information module used to store the General study information and Patient study information such as Study Instance UID, Study Date, Study Time, Referring Physician's Name, Study ID,.. etc. The Series information module is used to store the General series information, Frame of reference information, general equipment information such as Modality, Series Instance UID, Series Number, Conversion Type, Modality,.. etc. The Instance information used to store the image information, image pixel information, image plane information, SOP VOI LUT information such as Instance Number, Patient Orientation, Content Date, Content Time, Samples per Pixel, Bits Allocated, Pixel Representation, Planar Configuration, SOP Class UID, SOP Instance UID, Specific Character Set, etc.. The part of the DICOM file is also specified in the form of a data set which contains series of tags and nested child tags. These tags help carry information about the SOP instance such as the study, the series, the patient that it belongs to as well as other details regarding the image such as image pixel data, scan position data, etc. The Study, series and patient information are mostly used to index the DICOM Image in most PACS systems for faster retrieval of data. Following is the synopsis of the overall file structure and individual DICOM elements includes the tag, the associated information are part of the whole structure.

The Transfer syntax and its role in the DICOM standard enable devices to transfer information with each other even if they are running on different operating systems. Different operating systems and medical devices follow different standards and formats for data storage such as byte ordering when they store binary data. Compression methods help to transfer the large DICOM data over network. The VR encoding, Byte ordering and the compression techniques must first be understood and agreed upon to ensure that the two DICOM systems exchanging information understand each other during any communication between them. The transfer syntax is a set of encoding rules that helps specify these criteria through the use of UIDs. The Widely used Transfer techniques are shown in Table 1. The UID column denotes the UID value of the transfer syntaxes.

UID Values:

Table1. UID Values

Transfer Syntax	UID
Implicit VR Little-endian	1.2.840.10008.1.2
Explicit VR Little-endian	1.2.840.10008.1.2.1
Explicit VR Big-endian	1.2.840.10008.1.2.2
JPEG Lossless	1.2.840.10008.1.2.4.57

B. DICOM Data Model

The figure 3 shows that the DICOM consists of three major parts File header, Meta Information header, and Data object. The file header, consists of a file preamble (128-byte) followed by a prefix (4-byte) (mostly prefixed with 'DICM').The file preamble is used to ensure compatibility and consistency. The file preamble shall be skipped while parsing the DICOM file. The File meta-information header follows immediately after the file header and consists of a series of tags called "DICOM Elements". DICOM Elements are data set which specifies details such as the transfer syntax as well as other information regarding the device or implementation that created for receiving model. The third and last part of the DICOM files, is the data object.

In DICOM elements, '(0008, 0070)' indicates a tag belonging to group number of 0008 with an attribute number of 0070, the 'LO' denotes the data type or the Value Representation (VR), 'GE' is the actual value of the tag, '#2' helps specify the length of the value of GE. DICOM always encodes data using an even number of characters for text so an extra padding character is used for odd numbers. 1 represents the value multiplicity here, and 'Manufacturer' is the actual tag name as specified within the DICOM dictionary. Thus DICOM Element consists of attribute number, VR, value, multiplicity value and tag name. Dictionary implicitly defines the VR associated with each tag. In Implicit transfer syntax, VR is omitted and in the explicit transfer syntax, the VR is specified along with the tag. Explicit transfer is useful in DICOM information interchange.

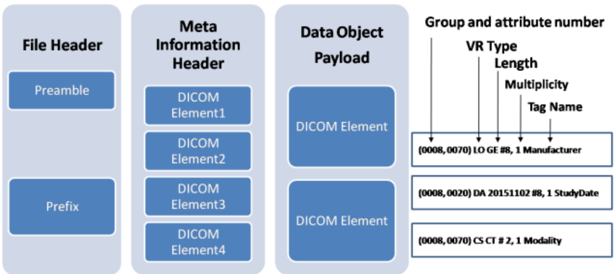

Fig. 3 DICOM File Structure

SCP is defined as Service Class Provider which provides the service. SCU is defined as Service Class User which uses the service. SOP is defined as Service Object Pairs which consists of both Service classes and objects. The abstract definition of an SOP is defined by UID (Unique Identifiers). The exchanges of commands during the network operations are called DIMSEs.

The paper organized as follows section 1 Introduction about DICOM, section 2 Data mining approaches in DICOM and finally the paper concluded in section 3.

II. DATA MINING APPROACHES IN DICOM

Before Cloud era, accumulating Big Data across a geographical location is not an easier task. But it's vital for medical informatics. Cloud paradigm is evolving in the health care domain. Readymade architecture, instant implementations are the core advantages of cloud Paradigm. But Cloud implementation is not a piece of cake. It requires more safety standards to evolve in the feature. We will discuss about this in the upcoming sections.

DOI 10.17148/IJARCCE.2016.53179

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

Data mining provides the intelligent base for medical informatics [4]. The objective of implementing Data mining is to discover the interesting, meaningful and actionable patterns hidden in large DICOM datasets. Thus Data mining helps healthcare organizations, health informatics, patient care, Patient assistive technology, Big data image analysis to extract critical data and automatic identification of unknown classes for further research and knowledge discovery. Treatment quality and survival rate is also getting increased by applying Data Mining Techniques.Data mining evolution helps the identification of grouping the patients related to several criteria and helps researchers to deploy new standards.

Thus combing distributed cloud paradigm and data mining, DICOM gains new strengths in solving age old problems. Each domain is evolving, so quality approach is required to implement Data mining in distributed DICOM. When compare to other Data collective domains, the calculative understanding of estimation and formation of hypothesis in DICOM is different altogether [3]. The Medical fields prioritize Individual patient care than research. The technical aspect should follow medical standards like HIPPA. Thus mining and distributed storage in DICOM are complex activities. The Core problems in applying data mining and distributed storage are briefly classified into technical and social. Technical aspects include homogeneity of DICOM and Social aspects include DICOM data Confidentiality.

The major areas of **Homogeneity of DICOM** are Volume and complexity [5]. The DICOM mining becomes highly complex since the interpretation of DICOM varies from radiologist to radiologist regarding the observations. The tests and diagnosis are different terms in DICOM. The medical condition is determined by different tests and accumulating the different test results leads to diagnosis. Sensitivity and specificity analysis requires for both Diagnosis and test. DICOM cannot be converted to raw mathematical formulas and models. It involves different expert opinion and analysis of DICOM. Hence the consistency is difficult to achieve. So Prognosis, Diagnosis and patient medical history and expert opinion plays critical part in mining DICOM.

Confidentiality issues is sharing and mining DICOM data raises many concerns including social, ethical and legal. Lena Griebel and Team raised the concerns in storing the DICOM patient data in a cloud environment [31]. Unsecured DICOM data shall be used by Organ trade groups, So security is a critical factor in Distributed storage. To overcome confidentiality of DICOM data, [32] Blackledge, J proposed a new encryption techniques named as stegacryption. Stegacryption will remove the confidential data in original image and then the patient data will be encrypted and again inserted into the same DICOM image. The disadvantages of this method are Increase in DICOM image size (DICOM by nature is a big data. Adding more data will affect DICOM distribution) and time criticality. Encryption and decryption will increase the reporting time. Since DICOM reporting involves Patient health, it's highly complex to encrypt and decrypt DICOM every time. As of now, cloud shall be used for elasticity, pay-per-use and broad network access related purpose. Cloud technologies in DICOM require more tools and standards to evolve in the future.

Bharat Rao and team stress the importance for large scale adoption of the data mining techniques in DICOM in the practical systems [49]. The technology is successful not by providing accuracy, but by assisting the radiologists and patients. Their team applied SVM initially and then moved to boosting algorithm and neural network. Since SVM is not specific to data domain and the key data characteristics are more generic. Neural network and boosting algorithm helps to evaluate which assumptions are appropriate for different kind of problems in data domain. The advantage of neural network and boosting algorithms are they allow mathematical tractability. The data is identically distributed and is never independent. By collaborating radiologist and data mining experts opinion the system shall be improved further.

A. Classification

Classification is the most commonly applied data mining technique. But it requires set of pre-classified examples to create a model. Based on the model the classification shall be applied on large dataset.Credit risk and fraud detection is identified using classification. Learning and classification are closed related to each other. Based on the history of patient diagnosis models will be create as a base for classification. P. Madheswaran, Rajendran, M., Naganandhini K., proposes image mining approach for medical image classification with multiple pre processing techniques [5].

Interestingly J. Umamaheswari, Dr. G. Radhamani, approaches data mining with a Hybrid approach for Classification of DICOM which consists of Multi Linear Discriminate Analysis (MLDA) and Support Vector Machine (SVM) [2]. Ref[1,2] jointly presented this technique. Here they are discussing about how to store the dicom images in distributed storage and how classification shall be applied to dicom image. As a result of this the image shall be queried quickly and more efficiently. This approach briefly consists of feature extraction and classification. Classification is based on the Gray Level Co-occurrence Matrix (GLCM) and histogram texture feature extraction method. The feature is selected using fuzzy rough set and Genetic Algorithm (GA). The proposed approach has high approximation capability and much faster convergence.

Support vector machines are supervised learning models with associated learning algorithms that analyse data used for regression analysis and classification. A Support Vector Machine (SVM) performs classification by finding the hyper plane which maximizes the margin between the two classes. The vectors (cases) that define the hyper plane are the support vectors.

A support vector machine applies classification and regression techniques to implement associated learning algorithms which analyse DICOM and recognize patterns [4]. The SVM works on hyper plane or set of hyper planes. SVMs are helpful in solving problems in bio-informatics,

DOI 10.17148/IJARCCE.2016.53179

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

particularly useful in analysing microarray expression data and detecting remote protein homologies. The SVM modelling error can be by limiting the model complexity. The model complexity shall be limited by applying the Structural risk minimization principle and VC theory (The Vapnik-Chervonenkis). This is useful in training the dataset. The Standard quadratic programming tools shall be used to solve optimization problems. The SVM Decision function is written as,

$$f(x) = \text{sign}\ (\sum_{i=1}^{N} y_i a_i k(\bar{x}, \bar{x}_i) + b)$$

Basically SVM is a binary classifier. The objective of SVM is to classify two classes of instances by finding the maximum separating hyper plane between two [33]. In order to allow more classes multiple methods are used. One-vsone method creates one binary class for each pair of classes. If there are three classes, then three binary classifiers will be created. In the simple form, Support vector machines are called as linear classifiers. Kernel trick shall be used to create non linear SVM by increasing the dimensionality of feature space. SVM uses many kernels. Most important kernels are Linear kernel, Polynomial kernel, Gaussian kernel. Gaussian kernel is special case for Radial Base Function. In the standard case, the distance used is the Euclidean distance. In the RBF kernel, the parameters determine, the width of the kernel, and d(x, y) is the distance metric.

In Machine learning Applications, SVM offers more robust and accurate methods among all well known algorithms. Since SVM has strong theoretical foundation, it requires only a dozen examples for training [38]. Strong community drives SVM development at good pace. Thus many efficient methods for training are evolving regularly. For binary classes, SVM finds the best classification function to distinguish between members of two classes in training data. For a linearly separable dataset, the two classes are separated by a linear classification function which passes through the middle of the two classes, separating the two in hyper plane f(x). After this, by testing the sign of the function f (xn), The new data instance xn can be classified, where xn belongs to the positive class for f (xn) > 0. By maximizing the margin between the two classes, the best function is found. The margin defined as the amount of space, or separation between the two classes. In Geometry, the shortest distance between the closest data points to a point on the hyper plane is defined as margin. This helps to define which hyper planes are qualified and which are not qualified. There are an infinite number of hyper planes, so only a few will qualify for SVM.

The maximum margin in hyper plane offers the best generalization ability. This also provides best classification performance on the training data. An SVM classifier attempts to maximize the function with respect to w and b: L P = 1 2 w−t, i=1 αi yi(xi + b) +t i=1 αi (2) where t is the number of training examples, and αi,i = 1,..., t, are positive numbers such that the L P derivatives with respect to αi are zero. αi = Lagrange multipliers and L P = Lagrangian. In the above equation, the hyper plane is defined by the vectors w and constant b.

A **Neural network** is a parallel distributed processor that has a propensity for storing experiential knowledge and making it available for users (Rumelhart a.o). Neural computing is the study of networks of adaptable nodes which store experiential knowledge using learning process. A neural network is a finite-state machine made up of elementary units called neurons (Minsky).

Darsana and team discusses about applying **Neural network** for image retrieval, fast computations **[17]**. By using fuzzy c-means algorithm the Image retrieval problem is solved. Darsana.B., Dr.G.Jagajothi, proposes neural network classifications for image retrieval [7]. They have used feature extraction method, Gabor filters and training neural network, precision and recall methods. They implemented the proposed system using java based platform. The query image must be pre-processed and the output objects extracted from the input query image in multi level database. The images are coached in the neural network. This helps in effective querying of images. They achieved good F-measure values using this system. Ref [2] has proposed the new system with neural network for DICOM. They are using this technique to overcome the slow rate in data analysis. They are also suggesting new ways to improve the system.

Decision tree is a widely used data mining method. In decision theory, a decision tree is a graph of decisions and their possible consequences, represented in form of branches and nodes. A decision tree contains a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node.

Decision trees use recursive data partitioning to extract useful data. This will be applied on DICOM data segmentation. Thus Decision trees helps to improve reliable predictions. [8] Decision tree, genetic algorithm and k-means algorithms are taken for their technical survey to find the biomedical image classifications. The objective of their study is to find the performance algorithms based on the classification accuracy in biomedical images. Image classification use to find expedient information from large amount of data. The Decision tree method shows 86 % accuracy, and genetic algorithm shows 91.53 % accuracy and k-means method shows 93 % accuracy in different biomedical images such as retinal images, x-ray images and CT images. From this k-means algorithm shows highest accuracy when compare with other two methods. They are suggesting using k-means algorithm for mining.

B. Association rule learning

Association rule learning is a method for finding the interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness.

Associative classification evolved by combining associative rule and classification methodology. Apriority and FP Growth algorithms used to generate class association rule. This helps to identify interesting correlations, frequent patterns, associations or casual structures in distributed DICOM Dataset. Rajesh Natarajan [4] proposed Relatedness-based data driven Approach to determination of Interestingness of association rules. They are discussing Item-relatedness, a composite of these relationships, which can help to rank interestingness of an AR. The approach presented, is intuitive and can complement and enhance classical objective measures of interestingness. They also work on TR beyond item-pairs, incorporating directional aspects of implications into interestingness evaluation, and testing on real dataset.

C. Clustering

Clustering is a process of partitioning a data set into clusters which contains set of meaningful sub-classes. Clustering helps users to understand the natural grouping or structure in a data set. Clustering is used either as a stand-alone tool to get insight into data distribution or as a pre-processing step for other algorithms. It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in pattern, machine learning, image analysis, information retrieval, and bioinformatics.

Clustering used to identify similar dicom objects. This is used for unknown datasets in DICOM. This provides a base to find a distributed pattern and correlation among DICOM attributes. Thus it helps to create groups automatically based on the patient data. Researches shall use this group to identify hidden information. [3] A. Mahendiran, N. Saravanan, N. Venkata Subramanian and N. Sairam, implements K-Means Clustering in Cloud environment. They deployed **K-Means algorithm** in Google Cloud using Google App Engine and Cloud SQL. [9]They used the segmentation of brain images using K-means and FCM. This study compared the efficiency of K-means and FCM for clustering MRI images. The segmentation of images using K-means is better than FCM for this dataset. Their work is the initial step for developing a system for information retrieval using data mining techniques. Clustering is pre-treatment part of other algorithms or kind of independent tool used to achieve data distribution, and can be used to determine isolated points [39]. CURE, KMEANS, DBSCAN, and BIRCH are the commonly used Clustering Algorithms. Every clustering method has respective advantages like: KMEANS is simple and easy to understand, DBSCAN is capable of filtering noises magnificently, and CURE method lacks input. Thus Priti and team highlight the importance to improve the new techniques in clustering.

K-means clustering uses vector quantization from signal processing. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. This cluster is used as the prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. This method produces exactly k different clusters of greatest possible distinction. The objective of K-Means clustering is to minimize total intra-cluster variance, or, the squared error function.

For small datasets, K-Means Algorithm provides good platform for mining. K-Means partitions n observations into 'k' clusters [39]. Within the cluster, the nearest mean is inspected. In K-Means k = number of clusters needed, A case is allocated to the cluster whose distance to the cluster mean is the insignificant. The algorithm is based on finding the k-means [5]. Using simple iterative method, K-means partitions the given dataset into number of clusters k(specified by user). The K-means algorithm operates on a set of D-dimensional vectors, D = {xi | i = 1,..., N}, where xi ∈ d = ith data point. The k points are picked in D is called centroids. The demerit of K-means is how to resolve quantify "closest" in the assignment.

D. Indexing

In indexing the meta data in DICOM is indexed just like database indexing. Indexing improves the meta data fetching process.Milton Santos and Team, proposes DICOM metadata access with consolidated view of departmental production data regardless of the heterogeneous imaging sources and subsidiary information systems [6]. Miton Santos and Team analysed this approach in a real time medical data set on 20 million images belonging to more than 467 thousand studies performed on more than 162 thousand patients [6]. This retrospective metadata study performed on three real time health care institutions. They implemented indexing system using Dicoogle application (Peer to peer Open source DICOM System) [6]. The proposed system handles multi model DICOM independent of device manufactures. Dicoogle implements DICOM metadata indexing. This leads to Data mining initiates for DICOM metadata. Multi model medical imaging modalities meta data are huge complex data source for mining. The team extracts the meta data using Dicoogle and answers multiple high level queries like "How many Patients have exams to the TORAX with Exposure time between X and Y". Without mining it's highly complex to answer the above question. They also analysed the global image statistics. Dicoogle provides visualization of global statistics and CSV data export provisions. This approach is helpful for analysing only the critical data which is required at the moment without traversing through all the images for each study for each patient.

The intensive survey has been conducted in different algorithms in data mining for DICOM data sets. The outcome of the survey produces the comparison of various algorithms based on data mining algorithm used experimental data set used outcomes of their research and demerits are listed in table 2.

III. CONCLUSION

The above survey focused on analysing and exploring more about distributed storage and data mining in DICOM and pros and cons of different Data mining algorithms used in DICOM. Various technical aspects help to improve DICOM, but still lot more security standards need to evolve in future. Applying Distributed storage and Data mining helps to save millions of life and cure millions of people in quality time. In classification techniques, SVM

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

is growing in a rapid pace. Associate laws have some demerits but still shall be used for mining meta data for small datasets. Complex datasets shall be mined using K-Means algorithm. For unknown meta datasets, clustering techniques shall be used effectively. For defined meta datasets, classification techniques shall be used. Thus by applying multiple mining techniques based on the dataset complexity, the data retrieval methods shall be improved. This will increase the survival rate of the patients. Lot more work need to be done on identifying the correct data mining techniques and distributed storage with safety standards. Mining techniques are helpful in predicting critical diseases at earlier stage. Billions of DICOM image archives are mined using different mining techniques to classify DICOM meta data. Clustering helps DICOM data to be trained easily and shall be use as base for future researches. My research work will be on selecting and apply best data mining techniques to retrieve image and meta data as quickly as possible.

Table 2 Comparison of Data mining Algorithm in DICOM

Author Name	Data Mining Algorithm	Data set	Outcomes	Demerits
S. Suganthira Thamilselvan. P Dr. J. G. R. Sathiaseelan M. Lakshmiprabha, 2015 [8]	Decision tree	X-ray	Classification Accuracy 86%	Sometime Accuracy will be loss
	Genetic algorithm	CT	91.53%	Difficult to find Classification Accuracy
	K-means method	Color retinal images	93%	Algorithms fails in Large datasets
M. Madheswaran, D. Anto Sahaya Dhas, 2015 [42]	SVM	Brain MRI Images	Classification Accuracy (98.83 %)	Low overall perform ance of the filter
Priti Sahu, Prof. Sunita Gond,2015 [47]	SVM	MRI	Classification Accuracy (97%)	More generic, Not specific for different data domain
B. Darsana, Dr. G. Jagajothi, 2014 [17]	Neural Network Classification	DICOM Dataset	Classification Accuracy (48.45%)	Low Retrieval rate
Rajkumar et al., 2010 [45]	Navie Bayes KNN Decision tree	ECG	Accuracy Rate 52.22%,52%,45	Low accuracy Rate
Reza Zare, M., Mueen, A. , Awedh, M. , Chaw Seng, W, 2013 [46]	Classification	X-RAY image	Classification Accuracy 92.5%.	Less training data set
Shweta Kharya, 2012, [50]	Decision tree	Mammograp hic Image	Classification Accuracy 93.62%	Iterative training producer, over training sensitive, need pruning
A. Kannan, V. Mohan, N. Anbazhagan, 2012, [24]	KNN,SVM	MRI	Classification Accuracy 97.3%	Low error rate
Jicksy Susan Jose, R. Sivakami, N. Uma Maheswari, R.Venkatesh, 2012 [29]	Association Rules	DICOM	Classification Accuracy 92%	Improved to obtain more representative features

REFERENCES

[1] Dwight A. Simon,"DICOM basics reference Basic DICOM Concepts with Healthcare Workflow," DICOM 2005 International Conference Budapest, Hungary ,September 2005.
[2] J. Umamaheswari, Dr. G. Radhamani, "A Hybrid Approach for Classification of DICOM Image," World of Computer Science and Information Technology Journal , 2011
[3] Krzysztof J. Ciosa, G. William Moore "Uniqueness of medical data mining," Artificial Intelligence in Medicine, March 2002
[4] Divdeep Singh Sukhpreet Kaur, "Scope of Data Mining in Medicine," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 6, June 2014
[5] Dhruv Madan Gopal, Aditya R, C Vishnu Kumar Reddy, Gautham S, Nagarathna N, "A Survey on Data Mining Applications, Techniques and Challenges in Healthcare," International Journal of Emerging Technologies and Innovative Research, April 2015, Volume 2, Issue 4
[6] J Westberg, S Krogh, C Brink, and I R Vogelius,"A DICOM based radiotherapy plan database for research collaboration and reporting," International Conference on the Use of Computers in Radiation Therapy (ICCR 2013)
[7] Steve G. Langer," A Flexible Database Architecture for Mining DICOM Objects: the DICOM Data Warehouse," Society for Imaging Informatics in Medicine, Published online: 12 November 2011,
[8] S. Suganthira Thamilselvan. P Dr. J. G. R. Sathiaseelan M. Lakshmiprabha, "A Technical Study on Biomedical image Classification using Mining Algorithms," March 2015
[9] Jeffrey D. Robinson,"Beyond the DICOM Header:Additional Issues in De identification," American Journal of Roentgenology, December 2014,Volume 6,Number 3
[10] Siji T.Mathew, Nachamai M,"Clustering of brain MRI image using data mining algorithm," International Conference, 1st February

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

2015, Pune, India, ISBN: 978-93-84209-85-8.

[11] Prasanna Desikan, Kuo-Wei Hsu, Jaideep Srivastava,"Data mining for healthcare management," International Conference on data mining, April 2011.

[12] Ximing Wang , BrentJLiu , ClarisaMartinez , XuejunZhang , Carolee JWinstein, "Development of an novel imaging informatics-based system with an intelligent work flow engine(IWEIS) to support imaging-based clinical trials," Comput Biol Med. 2016 Feb 1;69:261-9. doi: 10.1016/j.compbiomed.2015.03.024

[13] A. Mahendiran, N. Saravanan, N. Venkata Subramanian, N. Sairam, "Implementation of K-Means Clustering in Cloud Computing Environment ," Research Journal of Applied Sciences, Engineering and Technology 4(10): 1391-1394, 2012.

[14] Rajesh Natarajan, B. Shekar "Data mining (DM): poster papers: Relatedness-based data driven Approach to determination of Interestingness of association rules," Proceedings of the 2005 ACM symposium on applied computing SAC '05 ,March 2005.

[15] P. Madheswaran, Rajendran, M., Naganandhini, K., "An improved pre-processing technique with image mining approach for the medical image classification," Computing Communication and Networking Technologies (ICCCNT), 2010 International Conference , 29-31 July2010

[16] Milton Santos, Luis Bastiao, Nuno Neves, Dulce Francisco, Augusto Silva, nelson Pacheco Rocha., "Dicom metadata access, consolidation and usage in radiology department performance analysis. A non-proprietary approach," Conference on Enterprise Information System/International Conference on project Management/Conference on Health and social care Information system and Technologies, CENTERIS/ProjMan/Hcist 2015 October 7-9, 2015

[17] Darsana.B., Dr.G.Jagajothi,"An Efficient DICOM Image Retrieval method based on features and neural network classification," Journal of Theoretical and Applied Information Technology, 31st October 2014

[18] Praveen Pappula, Ramesh Javvaji, Rama B, "Experimental Survey on Data Mining Techniques for Association rule mining," International Journal of Advanced Research in Computer Science and Software Engineering, February 2014

[19] William Kovacs, Samuel Weisenthal, Les Folio,Qiaoyi Li, Ronald M. Summers, Jianhua Yao, "Retrieval, visualization, and mining of large radiation dosage data," Springer Science+Business Media New York (outside the USA) 2015.

[20] Tobias Dahlberg," Distributed Storage and Processing of Image Data", dec 2012.

[21] Dr. Guozhu Dong,Dong G, " Biomedical Data Mining," Editorial for International Journal of Biomedical Data Mining, Feb 2014.

[22] Mehraj Ali. U, John Sanjeev Kumar. A," Enhancement of Map Function Image Processing System Using DHRF Algorithm on Big Data in the Private Cloud Tool," International Journal of Innovative Research in Advanced Engineering (IJIRAE) ,ISSN: 2349-2163 Volume 1 Issue 5 (June 2014).

[23] Mahmoud Ismail,James Philbin," Fast processing of digital imaging and communications in medicine (DICOM) metadata using multi series DICOM format," Journal of Medical Imaging ,(Apr–Jun 2015).

[24] A. Kannan, V. Mohan, N. Anbazhagan , " MR Images Classification Using Hybrid KNNSVM Algorithm,"Signal Processing Research (SPR) Volume 1 Issue 1, December 2012

[25] Mihaela Gheorghe and Ruxandra Petre,"The importance of normalization methods for mining medical data," International journal of computers & technology Vol.14, No.8, June2014

[26] Richard Chen MD, Pattanasak Mongkolwat and David Channin,"Radiology Data Mining Applications using Imaging Informatics," InTech 2008

[27] Rajesh Natarajan, B. Shekar," A relatedness-based data-driven approach to determination of interestingness of association rules," Conference Paper DOI: 10.1145/1066677.1066803, January 2005

[28] Simón J. Rascovsky, Jorge A. Delgado, Alexander Sanz, Víctor D. Calvo, Gabriel Castrillón,"Informatics in Radiology Use of CouchDB for Document-based Storage of DICOM Objects," Presented as an education exhibit at the 2010 RSNA Annual Meeting,January 11, 2012.

[29] Jicksy Susan Jose, R.Sivakami, N. Uma Maheswari, R.Venkatesh, "An Efficient Diagnosis of Kidney Images Using Association Rules," International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 2, Issue 2,2012

[30] Steve G. Langer1,"DICOM Data Warehouse: Part 2", Society for Imaging Informatics in Medicine 2015

[31] J, SA & SFI, DIT, Dublin Ireland Al-Rawi, A.,Tobin, "Stegacryption of DICOM Metadata Black ledge," Irish Signals & Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies (ISSC 2014/CIICT 2014

[32] http://www.h17.org/fhir/media-example-dicom.json .html

[33] Smitha P, Shaji L, Dr. Mini M. G, "A Review of Medical Image Classification Techniques ," International Conference on VLSI, Communication & Instrumentation (ICVCI) 2011 Proceedings published by International Journal of Computer Applications,2011

[34] Richard Guindon, "DICOM Basics - Making Sense of the DICOM File Using Java and the Pixel Med DICOMToolkit", www.saravanansubramanian.com/Saravanan/Articles_On_Software/Entries/2014/9/28/DICOM_Basics_Making_Sense_of_the_DICOM_File.html

[35] Maria-Luiza Antonie , Osmar R. Za¨ıane , Alexandru Coman ,"Application of Data Mining Techniques for Medical Image Classification," Proceeding of the second International Workshop on Multimedia Data Mining, In conjunction with ACM SIGKDD conference san francisо ,2001

[36] "Novel Fuzzy Association Rule Image Mining Algorithm for Medical Decision Support System,"2010 ,International Journal of Computer Applications (0975 8887)Volume 1–No20

[37] Xindong Wu,Vipin Kumar,J.Ross Quinlan, Joydeep Ghosh , Qiang Yang , Hiroshi Motoda , Geoffrey J. McLachlan , Angus Ng , Bing Liu , Philip S. Yu , Zhi-Hua Zhou , Michael Steinbach , David J. Hand, Dan Steinberg "Top 10 algorithms in data mining " Springer-Verlag London Limited 2007, December 2007

[38] S.Nithya, G.ShineLet, " Bio-Medical Image Retrieval Using SVM," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 10, December 2012

[39] Petra Perner IBaI, Arno-Nitzsche-Str, Leipzig, Germany,"Mining Knowledge in Medical Image Databases," Data Mining and Knowledge Discovery: Theory, Tools, and Technology II, 359 (April 6, 2000), DOI:10.1117/12.381752

[40] Lashari S. A. and Ibrahim R., "Comparative Analysis of Data Mining Techniques for Medical Data Classification," Proceedings of the International Conference on Computing and Informatics ICOCI 2013

[41] P Thamilselvana , Dr. J. G. R. Sathiaseelanb, I.J. "A Comparative Study of Data Mining Algorithms for Image Classification," Published Online June 2015 in MECS (http://www.mecs-press.net) DOI: 10.5815/ijeme.2015.02.0

[42] M. Madheswaran and D. Anto Sahaya Dhas, "Classification of Brain MRI images using support vector machine with various Kernels," Biomedical Research 2015

[43] A.Kavipriya, B.Gomathy, "Data Mining Applications in Medical Image Mining: An Analysis of Breast Cancer using Weighted Rule Mining and Classifiers," IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661, ISBN: 2278-8727Volume 8, Issue 4 (Jan. - Feb. 2013)

[44] M. J. Abdi and D. Giveki, , "Automatic detection of Erythemato - Squamous diseases using PSO–SVM based on association rules," Engineering Applications of Artificial Intelligence, vol. 26, (2013), pp. 603-608.

[45] A. Rajkumar and G. S. Reena, "Diagnosis of Heart Disease Using Datamining Algorithm," Global Journal of Computer Science and Technology, vol. 10, no. 10, (2010).

[46] Reza Zare, Kuala Lumpur,Mueen, A. , Awedh, M. , Chaw Seng, W. ,"Automatic classification of medical X-ray images: hybrid generative-discriminative approach" ,IET Image Processing (Volume 7,Issue 5), July 2013

[47] Priti Sahu, Prof. Sunita Gond,s "An Efficient Technique of MRI Image Classification using PSO SVM," International Journal of Scientific & Engineering Research, Volume 6, Issue 6June2015

[48] Pranita Balaji Kanade,Prof. P.P. Gumaste,"Brain Tumor Detection Using MRI Images," International journal of innovative research in electrical, electronics, instrumentation and control engineering 2015

[49] R. Bharat Rao, Glenn Fung, Balaji Krishnapuram, Jinbo Bi, Murat Dundar, Vikas Raykar, Shipeng Yu, Sriram Krishnan, Xiang Zhou, Arun Krishnan, Marcos Salganicoff, Luca Bogoni, Matthias Wolf, Anna Jerebko, Jonathan Stoeckel **"**Mining Medical Images," Image and Knowledge Management-CAD and Knowledge Solutions (IKM-CKS) Siemens Medical Solutions USA, Inc., 51 Valley Stream Parkway, Malvern, PA-19355, 2009

[50] Shweta Kharya, "Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease," International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012