International Journal on Emerging Technologies 11(2): 954-957(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Classification using Association Rule Mining over Mammogram Images: A Review

N. Parashar[1,2]
[1]Assistant Professor, Department of CEA GLA University, Mathura (Uttar Pradesh), India.
[2]Research Scholar, Department of Computer Science and Engineering & IT
Madhav Institute of Technology & Science, Gwalior (Madhya Pradesh), India.

(Corresponding author: N. Parashar)
(Received 27 January 2020, Revised 23 March 2020, Accepted 26 March 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The mammograms are inspected with utmost consideration in order to detect true cases with least false rate. Mammogram samples are categorized into categories namely normal, benign and malignant by extracting and selecting best features from the given image datasets. Association Rule Mining is one of the method that can be applied for the classification of images by using the concept of image mining over image datasets. The rules in the association rule mining models can be weighted to illustrate the significance of generated rules where all attributes in the antecedent part of the rules have been usually weighted equally. Due to the out surge of huge amount of medical images on a larger scale day to day the procedure of examining images has turn out to be a tedious and critical task. However, this paper aims to compare the various techniques used in image mining for classification of mammograms; also the present status of image mining using image data association is being studied, providing references for upcoming research in the field of image mining.

Keywords: Association Rules, Classification, Feature Extraction, Image Association Rule Mining, Image Mining, Mammogram.

Abbreviations: SIFT, Scale Invariant Feature Transform; CNN, Convolutional Neural Network; SVM, Support Vector Machine; KNN, K-Nearest Neighbour; AUC, Area Under Curve; GANs, Generative Adversarial Networks.

I. INTRODUCTION

In today's era due to the out surge of huge amount of medical images on a larger scale day to day the practice of examining images has turn out to be a tedious and critical task. The data used can be released and announced to general users by cautiously probing these image datasets. Thus, image analysis is about mining the rational data from images; largely from digital images with digital image processing approaches [1] that often distinguish pixels from a digital image grounded on their size or texture. Split image is cast-off to enlarge the area well-defined by pixels of specific features as defined by the pixels.

Mammography is considered as one of the most reliable technique for the identification of tissues as normal, benign and malignant. This technique provides high accuracy but at increased cost. Wide research has been done in the field of mammography still there is no such method that can be used widely for the classification of mammograms as it requires high accuracy results. Therefore, for higher accuracy *Image Mining System* are required in order to extract features and information from vast amount of medical images datasets [2]. The focus is on finding image patterns and building information from images within a vast collection of image datasets.

In comparison with image mining, for representing patterns in similar kind of images association rules can be built for successful mining operations over medical image datasets [5]. The most insightful features during the mining procedure over image datasets are used to determine the strongest association rules. Therefore, the algorithm for mining of image datasets becomes rapid due to the usage of feature selection for more sensitive features and feature extraction.

A necessary step in Image Mining is classification of the obtained patterns. Automatic deriving of suitable decision criteria for classification signifies an obstacle still difficult to be overcome. Thus, a technique for developing effective and accurate association rules need to be developed.

This paper analyses various techniques used in image mining for classification of mammogram images along with their strength and limitations which will be useful in the processing of mammograms and in the generation of effective image association rules.

II. BACKGROUND

A. Image Mining

Image Mining is a data mining technique that is applied on images for mining and retrieving information which is of importance from vast amount of image datasets [2]. Various image patterns and features can be extracted from vast amount of image datasets using image mining. The core difference between image mining and image processing technique is that, in image mining the features and information is extracted from vast amount of image data sets while image processing is restricted to finding and/or understanding particular patterns and /or features from a single image [11].

Initially, for improving the quality of images in the image dataset pre-processing is performed for removing unwanted distortion and for enhancing the features of an image for further processing [12, 17]. Feature extraction is performed for transforming and generating important features from an image. For unlocking important features various data mining techniques can be applied to image mining. Obtaining the relevant information needed for effective decision-making will be evaluated and interpreted.

B. Image Association Rule Mining

Agrawal *et al.*, (1993) introduced the concept of Association Rule Mining over market basket analysis [2]. Ordonez and Omiecinski (1998) introduced the concept of image mining as a new approach for data mining [4]. At the same time period Zaïane *et al.*, (1998) used this notion of data mining over multimedia image dataset [3]. For finding relevant patterns, rules and procedure from vast amount of image datasets using the notion of data mining, the technique of image association rule mining can be used [3, 10]. In past few years, association rules have establish their use to be incorporated into large image datasets [5]. Although the current association rule mining process has been taken far in comparison to its use in the field of data mining it certainly undoes a promising research direction and the extent of additional exploration in the arena of Image Association Rule Mining [6, 7].

The two foremost methods can be used for mining of image datasets. The former method proposes mining over a large pool of image datasets alone, while the subsequent method requires mining over an integrated pool of image datasets and alphanumeric data [4]. An example of a previous methodology is to find out if there is any better pattern for a particular tissue or between tissues that are not the same as studying a group of mammogram images in the image dataset. An example of the following methodology may be containing patient records and their medical image datasets. For finding important associations the patient records and image datasets can be observed together.

The Image Association Rule Mining technique is applied to the database containing mammogram images dataset. The texture features mined from the mammogram images are referred to as *items* while each representable feature of an image i.e. the attribute value of an image is considered as the *transaction record* [3, 10]. The classification using association rule mining is comprised of two phases out of which the first phase represents the training phase which discovers association rules among the extracted features from an image. In the second phase classification is performed over the image dataset based on the extracted strongest association rules. Accuracy measures like support, confidence and lift are used for classification of mammograms. Based on the scores obtained using the strongest association rules the mammograms are classified into normal, benign and malignant category.

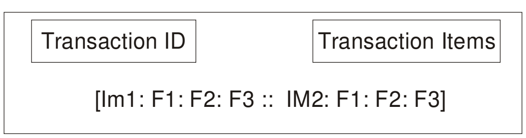

Fig. 1. Transaction Database.

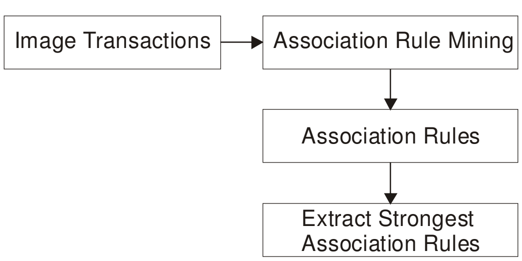

Fig. 2. Image Association Rule Mining.

Support is used to find how frequently the item appears in the database, while confidence is used for finding that how often the given rule is true for the given database.

Let A and B are the feature values of images then, a rule, A → B, holds and is measured as important if the support,

$$S = \frac{\text{Number of transactions containing AUB}}{\text{Total number of transactions}} \geq \text{MS}$$

where, MS is the minimum support assigned by the user and is used to find the frequent item sets in the image dataset and the confidence,

$$C = \frac{\text{No. of transactions which rule A} => B \text{ correctly}}{\text{Total No. of transactions in the data containing A}} \geq \text{MC}$$

where MC is the minimum confidence assigned by the user and is applied on the set of frequent items in the database in order to form association rules. The rule, A→B, infers that when a transaction t_k contains A, it most likely, also contains B.

The lift of a rule is defined as,

$$\text{Lift }(A \rightarrow Y) = \frac{\text{Support(AUB)}}{\text{Support(X)X Support(Y)}}$$

Association rule with lift=1 implies that occurrence probability of antecedent and consequent are independent of each other, and therefore no rule can be applied involving these two events.

Association rule with lift>1 finds the degree to which the two occurrences are dependent on one another, and potential rules are used for forecasting the consequent in forthcoming data set.

Association rule with lift<1, tells that the items are substitute to each other which means that the occurrence of one item has a negative effect on the presence of other item.

III. RELATED WORK

Ribeiro *et al.,* (2008) proposed PreSAGe and HiCARe algorithms that are used in a way where combining selection and subtraction are performed in one step in order to reduce the complexity of mining using PreSAGe while HiCARe is a new interactive entity with the important properties of assigning multiple keywords per image to enhance high-accuracy diagnostics. However the feature extraction step must be optimized for obtaining more representable features from the image datasets [9].

Deshmukh and Bhosle *et al.*, (2016) proposed the classification of mammogram images using SIFT by optimizing the SIFT features using PreARM algorithm and organization rules are augmented using an algorithm for multiple robustness functions. 93.75% and 0.932 classification accuracy and derivation of functional factor i.e. ROC is obtained using the proposed classifier over DDSM dataset. Better accuracy with least area under ROC curve value is provided by comparing the random forest path. Even after applying the association rules effectively over the images dataset the accuracy of the entire system is maintained. For reducing the number of complex rules and system complexities the algorithm proves to be very effective. Also, in addition this method is applicable over other medical image analysis programs [13].

Fathima Abubacker *et al.*, (2016) proposed an extended classifier for a neural network tested in a data set derived from the University of South Florida of DDSM. Statistical features are mined and genetic analysis is performed over the feature selection process using the multivariate filter. The program identifies functional data distinctions that the rules of input relations include formation and training using neural network and reveal higher levels of accuracy, precision, sensitivity, and precision. The properties of fuzzy parameters along with neural network objects are examined in this work and it is deduced that associative classifiers using fuzzy neural network (ACFNN) performs better when dealing with ambiguous information and classification accuracy up to 95%. The proposed method performs better as compared to traditional classification methods as it provides higher accuracy by deriving weights from the input rules of an input [14].

Dubrovina *et al*., (2016) proposes a new theoretical framework based on the study of tissue differentiation proving applicability in the classification of mammogram image segmentation. The suggested method uses CNN to learn aspects of auto-discrimination during student training. CNN has been trained to use the smart method; this ensures an ample amount of training samples. To accelerate class-pixel-smart prediction, compliant layers are used as an alternative of fully coupled ones. The classification accuracy is maintained by this method even after it produces at a faster rate of two magnitudes [15].

Deshmukh and Bhosle *et al*., (2016) have developed an image mining method for the issuance of strict and closely related association rules and materials used in digital millimeters. The retrieval of obsolete data through the use of the Apriori Algorithm was overcome by the newly proposed ESAR algorithm [16].

Deshmukh and Bhosle (2017) has made the best of the GLCM signals and assembly rules using the PreARM and ESAR algorithm respectively. The presented system is applicable to the MIAS data and DDSM mammogram data. In addition the mammogram image segmentation is done with the help of well-established organizational rules. This proposed method provides better accuracy for classifying approximately 92% of MIAS and 94% of DDSM databases compared to other methods of engagement. The area under the ROC curve value obtained with the proposed MIAS and DDSM database is 0.9656 and 0.9285, respectively [18].

Sonar *et al*., (2017) presented modified mammogram segmentation for improving the classification accuracy of hybrid KNN-SVM. MIAS and DDSM datasets were used for confirming the test results. KNN, SVM and Random Forest Classifier were used for comparing the test results and it was found that an accuracy of 9.18 % as obtained for normal/abnormal sections in the mammograms and 100% non-invasive segmentation. The feature set released looks steady as it delivers the same accuracy on the hybrid SVM-KNN and Random Forest Classifier. But at this point, it will not accomplish anything with the severity of the symptoms as they try to engage more and more students. However, by reducing the number of classifier training items the complexity of the proposed classifier can be reduced. With that feature implementation process using a modified domain method can be used [19].

Wu *et al*., (2018) uses a multi-dimensional GAN with a mask infill (ciGAN), and demonstrates that the proposed GAN is capable of extracting realistic lesions, enhancing the performance of subsequent sequences over traditional methods. The proposed strategies address critical issues in other GAN structures, such as training robustness and resolution data. Fears of detail and class imbalances are collective barriers to medical thinking, and thus strategies can help address these problems in a variety of settings [20].

Zeng and Gimenez *et al*., (2019) developed a model for mammography practitioners that helped in taking decisions related to classification. Based on a total of 112,433 mammographic cases from 36,111 patients and 13 radiologists at 2 different centers with 1.1% of fatal venous injury, the Bayesian network (BN) was trained to estimate the risk of wound injury. The sensitivity and specificity for each radiologist were compared with the BN model using their optical limit or the 2% standard limit suggested by the BI-RADS. The results showed a significant reduction in mammography screening of false positives with a significant increase in false positives [21].

IV. COMPARATIVE EVALUATION

The Table 1 provides the summary of various techniques and algorithms along with the accuracy ratio for the classification of mammogram image datasets in the healthcare sector.

Table 1: Summary of Techniques and Algorithms with accuracy ratio over mammogram datasets.

Algorithms/ Techniques	Hands on	Accuracy over Mammogram Image Dataset	References
Associative Rule based PreSAGe algorithm + HiCARe associative classifier	Mammogram Classification	92% accuracy and 95% high sensitivity.	[9]
SIFT based Associative Classifier	Mammogram Classification	93.75% accuracy and 0.932 ROC curve value.	[13]
Genetic Association Rule Miner + Neural Network	Mammogram Classification	95.1% accuracy.	[14]
CNN	Tissue Classification	Raw DNN output.	[15]
Extract Strong Association Rule Algorithm (ESAR)	Generation of strong association rules	252 strong and effective ARs obtained out of 591ARs.	[16]
GLCM based Associative Classifier	Classification using optimized association rules	92% accuracy for MIAS dataset and 94% accuracy for DDSM dataset.	[18]
SVM+KNN	Mammogram Classification	100% accuracy for DDSM database and 94% accuracy for MIAS database.	[19]
Generative Adversarial Networks (GANs)	Data Augmentation	AUC of 0.896 is achieved.	[20]
Bayesian Network	Reduction in false positive results	0.01% increase in false negative results and 28.9% reduction in false positive results.	[21]

V. CONCLUSION

Various techniques that have been used by previous researchers working in the process of mining are analysed in this paper. The focal point in which these methods are applied and its strengths and limitations to fulfil real-time needs while processing medical information is necessary to obtain and issue effective and efficient image association mining rules. Therefore, the methods discussed above are useful for the processing of mammograms and for capturing local information for images such as geometric features, textural features etc. Comparisons between images are made to determine whether they are compatible or not based on the rules of the relationship of the images provided. Also, in the efficient and accurate generation of associations using a variety of implementation strategies can work.

VI. FUTURE SCOPE

Based on extensive research we can conclude that a useful assessment of the effects of image mining should be applied to the functional boundaries. However, there is still a lot of empty space for upcoming development and new practices need to be established in an effective and accurate way that can assist in the decision-making process of several application forums.

Conflict of Interest. The author declares no conflict of interest.

REFERENCES

[1]. Gonzalez, R.,& Woods, R. (1992). Digital Image Processing, Second, Pearson Addison-Wesley Publications.

[2]. Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases, *in Proceedings of the ACM SIGMOD ICMD*, Washington DC, 207-216.

[3]. Zaïane, O. R., Han, J., Li, Z. N., & Hou, J. (1998). Mining multimedia data. In *Proceedings of the 1998 conference of the Centre for Advanced Studies on Collaborative research*, 1-18, IBM Press.

[4]. Ordonez, C., & Omiecinski, E. (1998). *Image mining: A new approach for data mining. Georgia Institute of Technology,* 1-22.

[5]. Ordonez, C., & Omiecinski, E. (1999). Discovering association rules based on image content. In *Proceedings IEEE Forum on Research and Technology Advances in Digital Libraries*, 38-49. IEEE.

[6]. Zaiane, O. R., Han, J., & Zhu, H. (2000). Mining recurrent items in multimedia with progressive resolution refinement. In *Proceedings of 16th International Conference on Data Engineering (Cat. No. 00CB37073)* (pp. 461-470). IEEE.

[7]. Olukunle, A., & Ehikioya, S. (2002). A fast algorithm for mining association rules in medical image data. In *IEEE CCECE2002. Canadian Conference on Electrical and Computer Engineering. Conference Proceedings (Cat. No. 02CH37373)*, *2*, 1181-1187. IEEE.

[8]. Tseng, V. S., Wang, M. H., & Su, J. H. (2005). A new method for image classification by using multilevel association rules. In *21st International Conference on Data Engineering Workshops (ICDEW'05).* IEEE.

[9]. Ribeiro, M. X., Traina, A. J. M., Traina, C., & Paulo, M. (2008). An Association Rule Based Method to Support Medical Image Diagnosis with Efficiency, *IEEE Transactions on Multimedia*, *10*(2), 277-285.

[10]. Han, J., & Kamber, M. (2012). Data Mining Concepts and Techniques, *Third. Norgan Kaufmann*, Elsevier.

[11]. Thirunavu, K., & Manikandaprabhu, P. (2013). Function and Information Driven Frameworks for Image Mining - A Review. *International Journal of Advanced Research in Computer and Communication Engineering*, *2*(11), 4202-4206.

[12]. Joseph, N., & Wilson, A. (2014). Retrieval of images using data mining techniques. *International Conference on Contemporary Computing and Informatics,* 204-208.

[13]. Deshmukh, J., & Bhosle, U. (2016). SIFT with associative classifier for mammogram classification. In *2016 International Conference on Signal and Information Processing (IConSIP)*, 1-5. IEEE.

[14]. Fathima Abubacker, N., Azman A., Doraisamy, S., & Murad, M. A. (2016). An integrated method of associative classification and neuro-fuzzy approach for effective mammographic classification. *Neural Computer and Applications, Springer*, 3967-3980.

[15]. Dubrovina, A., Kisilev, P., Ginsburg, B., Hashoul, S., & Kimmel, R. (2016). Computational mammography using deep neural networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *Taylor & Francis, 6*(3), 243-247.

[16]. Deshmukh, J., & Bhosle, U. (2016). Image Mining using Association Rule for Medical Image dataset, *in Proceedings of International Conference on Computational Modeling and Security, Elsevier*, *85,* 117-124.

[17]. Htake Khaung & Tin, H. (2017). Learning the Comparison of Image Mining Technique and Data Mining Technique, *in Proceedings of 4th International Conference on Multidisciplinary Research & Practice.*

[18]. Deshmukh, J., & Bhosle, U. (2017). GLCM based Improved Mammogram Classification using Associative Classifier. *I. J. Image, Graphics and Signal Processing*, *7*, 66-74.

[19]. Sonar, P., Bhosle, U., & Choudhury, C. (2017). Mammography Classification Using Modified Hybrid SVM-KNN. *International Conference on Signal Processing and Communication*, 305-311.

[20]. Wu, E., Wu, K., Cox, D., & Lotter, W. (2018). Conditional Infilling GANs for Data Augmentation in Mammogram Classification. *Lecture Notes in Computer Science,* 98-106.

[21]. Zeng, J., Gimenez, F. (2019). A Probabilistic Model to Support Radiologists Classification Decisions in Mammography Practice. *Medical Decision Making*, *39*(3), 208-216.

How to cite this article: Parashar, N. (2020). Classification using Association Rule Mining over Mammogram Images: A Review. *International Journal on Emerging Technologies*, *11*(2): 954–957.