

# Priority Based Apriori Algorithm For Cancer Prediction Using Fuzzy Classification

Dr. A. Padmapriya, M.C.A.,M.Phil.,Ph.D<sup>#1</sup>, K. Silamboli Chella Maragatham<sup>#2</sup>

<sup>#</sup>Department of Computer Science and Engineering, Alagappa University  
Karaikudi

**Abstract**— In this work, we take advantage of association rule mining to support the prediction of cancer. In machine learning and statistics, feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features for use in model construction. We reduce the features using FUZZY based rough set theory and then apply priority based approach. We propose priority based apriori for rule generation. Finally we apply the FUZZY classification approach to classify the dataset as normal or abnormal prediction of cancer.

## Keywords-

Featurereduction,associationrules,classification, prediction

## I. Introduction

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

This paper is focusing to develop a method based on association rule-mining to enhance the prediction of cancer. This is to implement a computer-aided decision support system for an automated diagnosis and classification of images. The method uses association rule mining to analyze the medical images and automatically generates suggestions of diagnosis. It combines automatically extracted low-level features from images with high-level knowledge given by a specialist in order to suggest the diagnosing.

The widely used and well-known data mining functionalities are Characterization and Discrimination, content based analysis, Association

Analysis, Categorization and Prediction, Outlier Analysis, Evolution Analysis [21].

CLASSIFICATION algorithms usually require an adequate and representative set of training data to generate an appropriate decision boundary among different classes. This requirement still holds even for ensemble (of classifiers)-based approaches that resample and reuse the training data. However, acquisition of such data for real-world applications is often expensive and time consuming. Hence, it is not uncommon for the entire data set to gradually become available in small batches over a period of time. In such settings, an existing classifier may need to learn the novel or supplementary information content in the new data without forgetting the previously acquired knowledge and without requiring access to previously seen data. The ability of a classifier to learn under these circumstances is commonly referred to as “incremental learning.” On the other hand, in many applications that call for automated decision making, it is not unusual to receive data obtained from different sources that may provide complementary information. A suitable combination of such information is known as “data” or “information” fusion, and can lead to improved accuracy of the classification decision compared to a decision based on any of the individual data sources alone. Consequently, both incremental learning and data fusion involve learning from different sets of data. If the consecutive data sets that later become available are obtained from different sources and/or consist of different features, the incremental learning problem turns into a data fusion problem. Recognizing this conceptual similarity, we propose an approach based on an ensemble of classifiers originally developed for incremental learning as an alternative and surprisingly well-performing approach to data fusion.

Different data mining techniques available to solve data mining problems are classification, association rule mining, time series analysis, clustering, summarization, sequence discovery. Out of these Association rule mining is popular and well researched data mining technique for discovering interesting relations between variables in large

databases. There are various association rule mining algorithms like Apriori, FP-Growth, Partition based algorithm, Incremental update, Haskell based approach, Fast algorithm and other Apriori based algorithms etc. These algorithms try to find out correlation or association among data in large volume of database. Most of the previous studies for frequent itemsets generation adopt an Apriori algorithm that has exponential complexity (high execution time). In this study, we propose an algorithm that will reduce execution time by means of generating itemsets progressively from static database namely Priority based Apriori.

## 2. RELATED WORK

### 2.1. Feature Extraction:

When dealing with medical images, the earliest phase of a CAD (computer aided diagnosis) system demands to extract the main image features regarding a specific criterion. Essentially, the most representative features vary according to the image type (e.g. mammogram, brain or lung) and according to the focus of the analysis (e.g. to distinguish nodules or to identify brain white matter). M.X. Ribeiro [18] defined as The Image Diagnosis Enhancement through Association rules (IDEA) method can work with various types of medical images, and with different focus of analysis. However, for each type of image and goal, an appropriate feature extractor should be employed.

### 2.2. Content based image retrieval:

The earliest use of the term content-based image retrieval in the literature seems to have been by Kato [1992], to describe his experiments into automatic retrieval of images from a database by colour and shape feature. The term has since been widely used to describe the process of retrieving desired images from a large collection on the basis of features (such as colour, texture and shape) that can be automatically extracted from the images themselves. The features used for retrieval can be either primitive or semantic, but the extraction process must be predominantly automatic. Retrieval of images by manually-assigned keywords is definitely not CBIR as the term is generally understood – even if the keywords describe image content. CBIR differs from classical information retrieval in that image databases are essentially unstructured, since digitized images consist purely of arrays of pixel intensities, with no inherent meaning. One of the key issues with any kind of image processing is the need to extract useful information from the raw data (such as recognizing the presence of particular shapes or textures) before

any kind of reasoning about the image's contents is possible. Image databases thus differ fundamentally from text databases, where the raw material (words stored as ASCII character strings) has already been logically structured by the author [Santini and Jain, 1997]. There is no equivalent of level 1 retrieval in a text database. CBIR draws many of its methods from the field of image processing and computer vision, and

is regarded by some as a subset of that field. It differs from these fields principally through its emphasis on the retrieval of images with desired characteristics from a collection of significant size. Image processing covers a much wider field, including image enhancement, compression, transmission, and interpretation. While there are grey areas (such as object recognition by feature analysis), the distinction between mainstream image analysis and

CBIR is usually fairly clear-cut. An example may make this clear. Many police forces now use automatic face recognition systems. Such systems may be used in one of two ways. Firstly, the image in front of the camera may be compared with a single individual's database record to verify his or her identity. In this case, only two images are matched, a process few observers would call CBIR. Secondly, the entire database may be searched to find the most closely matching images.

Large number of images is generated by hospitals and clinics every day. These images play very important role in diagnosis of diseases, medical research and education. CBIR systems have identified as an important research topic in radiology to facilitate diagnostic decision support for medical image interpretation using gradually increasing clinical data [1]. Li et al. [2] presents a CBIR-based tool to aid in radiological diagnosis. Kawata et al. [3] developed a CBIR system on lung nodule in 2004 considering shape descriptors and density histograms to retrieve 3-D lung nodules but precision and recall of this CBIR system was not reported. Lam et al. [4] in 2007 developed an open source pulmonary nodule image retrieval framework using Haralick features from grey-level co-occurrence matrix. Melanoma and non-melanoma skin cancers currently constitute one of the most common malignancies in the caucasian population, and the worldwide incidence and mortality rates are continuously increasing [5]. In particular melanoma incidence has increased more than any other cancer, reaching currently 18 new cases per 100,000 population per year in the United States [6]. Because advanced skin cancers remain incurable, early detection and surgical excision currently is the only approach to reduce mortality.

### 2.3. Current level CBIR techniques:

CBIR operates on a totally different principle, retrieving stored images from a collection by comparing features automatically extracted from the images themselves. The commonest features used are mathematical measures of colour, texture or shape; hence virtually all current CBIR systems.

### 2.4. Computer aided diagnosis:

The long range goal is to improve the accuracy and consistency of breast cancer diagnosis by developing a Computer Aided Diagnosis (CAD) system for early prediction of breast cancer from patients' mammographic findings and medical history. Specifically, this system will predict the malignancy of non-palpable lesions that are examined with diagnostic mammography and are considered for biopsy.

CAD is used in the diagnosis of breast cancer, lung cancer, colon cancer.

### 2.5. Breast cancer:

CAD is used in screening mammography (X-ray examination of the female breast). Screening mammography is used for the early detection of breast cancer. CAD is especially established in US and the Netherlands and is used in addition to human evaluation, usually by a radiologist. The first CAD system for mammography was developed in a research project at the University of Chicago. Today it is commercially offered by iCAD and Hologic. There are currently some non-commercial projects being developed, such as Ashita Project, a gradient-based screening software by Alan Hshieh, as well. However, while achieving high sensitivities, CAD systems tend to have very low specificity and the benefits of using CAD remain uncertain. Some studies suggest a positive impact on mammography screening programs,[7][8] but others show no improvement[9][10]. A 2008 systematic review on computer-aided detection in screening mammography concluded that CAD does not have a significant effect on cancer detection rate, but does undesirably increase recall rate (i.e. the rate of false positives). However, it noted considerable heterogeneity in the impact on recall rate across studies.[11] Procedures to evaluate mammography based on magnetic resonance imaging exist too.

### 2.6. Lung cancer (bronchial carcinoma)

In the diagnosis of lung cancer, computed tomography with special three-dimensional CAD systems are established and considered as gold standard. At this a volumetric dataset with up to 3,000 single images is prepared and analyzed. Round lesions (lung cancer, metastases and benign changes) from 1 mm are detectable. Today all well-known

vendors of medical systems offer corresponding solutions. Early detection of lung cancer is valuable. The 5-year-survival-rate of lung cancer has stagnated in the last 30 years and is now at approximately just 15%. Lung cancer takes more victims than breast cancer, prostate cancer and colon cancer together. This is due to the asymptomatic growth of this cancer. In the majority of cases it is too late for a successful therapy if the patient develops first symptoms (e.g. chronic croakiness or hemoptysis). But if the lung cancer is detected early (mostly by chance), there is a survival rate at 47% according to the American Cancer Society.[12] At the same time the standard x-ray-examination of the lung is the most frequently x-ray examination with a 50% share. Indeed the random detection of lung cancer in the early stage (stage 1) in the x-ray image is difficult. It is a fact that round lesions vary from 5–10 mm are easily overlooked.[13] The routine application of CAD Chest Systems may help to detect small changes without initial suspicion. Philips was the first vendor to present a CAD for early detection of round lung lesions on x-ray images.[14].

### 2.7. Colon cancer:

CAD is available for detection of colorectal polyps in the colon. Polyps are small growths that arise from the inner lining of the colon. CAD detects the polyps by identifying their characteristic "bump-like" shape. To avoid excessive false positives, CAD ignores the normal colon wall, including the haustral folds. In early clinical trials, CAD helped radiologists find more polyps in the colon than they found prior to using CAD.[15][16]

Feature Selection is a process that attempts to select a subset of features, satisfying a combination of application and methodology-dependent criteria: minimizing the cardinality of the feature subset; ensuring classification accuracy does not significantly decrease; and approximating the original class distribution with the class distribution given the selected features.

R. Agrawal et al. [17] present a method based on association rule-mining to enhance the diagnosis of medical images. It combines low-level features automatically extracted from images and high-level knowledge from specialists to search for patterns. The proposed method analyzes medical images and automatically generates suggestions of diagnosis employing of association rules. The suggestions of diagnosis are used to accelerate the image analysis performed by specialists as well as to provide them an alternative to work on.

## 3. PROPOSED METHOD

In this section we discuss the following methods

- Collect Dataset
- Feature selection
- Feature Reduction
- Rule Generation
- Classification
- Prediction

### 3.1. Collect dataset:

we present cancer dataset performed to validate the fuzzy based rough set theory to predict the cancer. But in existing method have only performed the diagnosis of images. In this paper I suggest that directly implement the input as the image value descriptions. The image value descriptions is denoted as cell size, cell shape, mitosis of given image for predicting the cancer.

### 3.2.Feature selection:

In feature selection method to proposed a fuzzy based rough set theory will be used. The main goal of this paper for feature selection to reducing the features size by using mathematical calculations. That calculation such as mahalnobis distance to combine the fuzzy based rough set theory to produce the reducing features.

Feature Selection (FS) or Attribute Reduction techniques are employed for dimensionality reduction and aim to select a subset of the original features of a data set which are rich in the most useful information. The benefits of employing FS techniques include improved data visualization and transparency, a reduction in training and utilization times and potentially, improved prediction performance. Many approaches based on rough set theory up to now, have employed the dependency function, which is based on lower approximations as an evaluation step in the FS process. However, by examining only that information which is considered to be certain and ignoring the boundary region, or region of uncertainty, much useful information is lost.

### 3.3.Feature Reduction:

In feature reduction we convert the weighted features into fuzzy values [0,1].the fuzzy values are set as 0-low,0.5-medium,1-high.In this paper to employ association rules to weight features according to their significance, promoting continuous feature selection to represent the image value description. Continuous feature selection techniques assign the features to apply the fuzzy values for each features, allowing the most important is highest features only to be selected for computation.

### 3.4.Rough set theory working method:

The work on rough set theory (RST) offers a formal methodology that can be employed to reduce the dimensionality of data sets, as a preprocessing step to assist any chosen modeling method for learning from data. It assists in identifying and selecting the most information-rich features in a data set. This is achieved without transforming the data,while simultaneously attempting to minimize information loss during the selection process. In terms of computational effort, this approach is highly efficient, and is based on simple set operations, which makes it suitable as a preprocessor for techniques that are much more complex.

### 3.5.Fuzzy based rough set approaches:

A fuzzy-rough set is defined by two fuzzy sets, fuzzy lower and upper approximations, obtained by extending the corresponding crisp rough set notions. In the crisp case, elements that belong to the lower approximation (i.e., have a membership of 1) are said to belong to the approximated set with absolute certainty. In the fuzzyrough case, elements may have a membership in the range [0, 1], allowing greater flexibility in handling uncertainty.

### 3.6.Association Rules:

Association Rule mining is the technique for knowledge discovery. It is a well-known method for discovering correlations between variables in large databases.A selected highest features only to applied the association rules.the association process rule manner is combination between the features .We propose priority based apriori for rule generation & apply FUZZY classification. After complete the training process get the testing set from the user. Compare testing set into training result and also apply classification technique for prediction. Finally we display predicted result for testing set.

### 3.7.Classification:

In classification techniques method to be classified into two classes. The output of the result will be present in below manner.

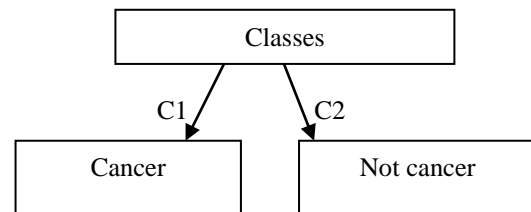


Figure1:Classification of two classes

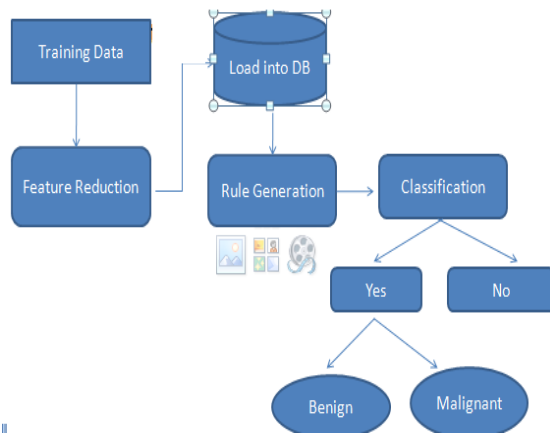


Figure 2:Flow diagram

3.8.Prediction:

Prediction is similar to that of data classification. However, for prediction, instead of using the term “class label attribute”, the attribute can be referred to simply as the predicted attribute.

We can predict whether the new user is affected by cancer or not and also updates the stages of cancer.

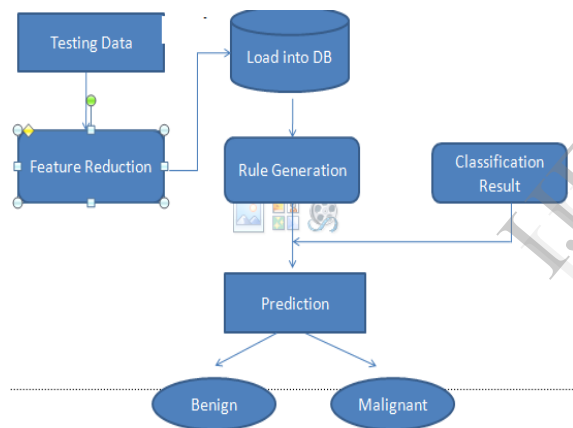


Figure 3:flow diagram of predicting cancer stages

We can predict whether the new user is affected by cancer or not and also updates the stages of cancer.

4 .Ilustration Example:

4.1 .STEP:1 Feature selection:

Fuzzy Based roughset theory using DMRSAR:

Fuzzy Based roughset theory method is combine the mahalnobis distance with Distance metric rough set attribute reduction algorithm to reduce the features.

DMRSAR algorithm in roughset theory performs the mahalnobis distance of continuous values.mahalnobis

distance processes a each feature separately and range of values in 4N

Steps.let f be a feature and fl be the value of the feature f in the image value descriptions as i.fl be the value with instance class label ci.we refer to an image instances Ii as the pair(fi,ci)

4.1.1.Standard Deviation Formulas

Let the features are representd by a set F={f1,f2,f3.....fp}

Let a features are 9, 2, 5, 4, 12, 7, 8, 11.

Mahalnobis distance formulas:

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^N \frac{(x_i - y_i)^2}{s_i^2}}$$

Prediction and classification also differ in the methods that are used to build their respective models. As with classification, the training set used to build a predictor should not be used to assess its accuracy. An independent test set should be used instead. the attribute can be referred to simply as the predicted attribute predictor is estimated by computing an error based on the difference between the predicted value and the actual known value of y for each of the test tuples, X.

To calculate the standard deviation of those numbers:

- 1.Work out the Mean (the simple average of the numbers)
2. Then for each number: subtract the Mean and square the result
3. Then work out the mean of those squared differences.
4. Take the square root of that value.

Let a features are 9, 2, 5, 4, 12, 7, 8, 11, 9, 3, 7, 4, 12, 5, 4, 10, 9, 6, 9, 4

Formulas for standard deviation

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

Step 1:Work out the mean

Feature of values: 9, 2, 5, 4, 12, 7, 8, 11, 9, 3, 7, 4, 12, 5, 4, 10, 9, 6, 9, 4

The mean is  $(9+2+5+4+12+7+8+11+9+3+7+4+12+5+4+10+9+6+9+4) / 20 = 140/20 = 7$

So:  $\mu = 7$

**Step 2:**Then for each number: subtract the Mean and square the result

This is the part of the formula that is:

$$(x_i - \mu)^2$$

So what is  $x_i$  ? They are the individual x values 9, 2, 5, 4, 12, 7, etc...

In other words  $x_1 = 9, x_2 = 2, x_3 = 5, \text{ etc.}$

So it "for each value, subtract the mean and square the result".

Example (continued):

$$(9 - 7)^2 = (2)^2 = 4$$

$$(2 - 7)^2 = (-5)^2 = 25$$

$$(5 - 7)^2 = (-2)^2 = 4$$

$$(4 - 7)^2 = (-3)^2 = 9$$

$$(12 - 7)^2 = (5)^2 = 25$$

$$(7 - 7)^2 = (0)^2 = 0$$

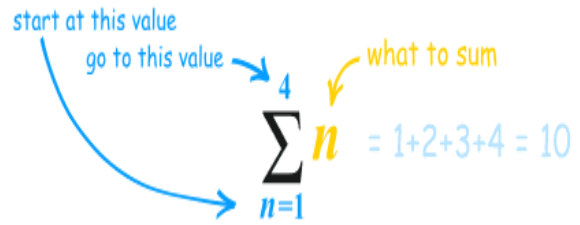
$$(8 - 7)^2 = (1)^2 = 1$$

... etc ...

**Step 3:**Then work out the mean of those squared differences.To work out the mean, add up all the values then divide by how many.First add up all the values from the previous step.

Add them all up" in mathematics? We use "Sigma":  $\Sigma$

Sigma\_Notation means to sum up as many terms as



Sigma Notation

We want to add up all the values from 1 to N, where  $N=20$

In our case because there are 20 values:

**Example (continued):**

$$\sum_{i=1}^N (x_i - \mu)^2$$

Which means: Sum all values from  $(x_1-7)^2$  to  $(x_N-7)^2$ .

We already calculated  $(x_1-7)^2=4$  etc. in the previous step, so just sum them up:

$$= 4+25+4+9+25+0+1+16+4+16+0+9+25+4+9+9+4+1+4+9 = 178$$

But that isn't the mean yet, we need to divide by how many, which is simply done by multiplying by "1/N":

Example (continued):

$$\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

Mean of squared differences =  $(1/20) \times 178 = 8.9$

(Note: this value is called the "Variance")

**Step 4.** Take the square root of that and you are done!

Example (concluded):

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

$$\sigma = \sqrt{8.9} = 2.983...$$

Finally to apply by Mahalanobis Distance formula,

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^N \frac{(x_i - y_i)^2}{s_i^2}}$$

Where  $(x,y) = \frac{\sqrt{(9-2)}}{2.9} = 49/2.9 = 1.6$  so for this example

answer is  $A=1.6$  and so etc..

- $A=8.5$
- $B=0.5$
- $C=0$

Apply Fuzzy values (0-low, 0.5-medium, 1-high)

- $A=8.5$ (highest value)
- $B=0.5$
- $C=0$

#### 4.2.STEP 2:Association Rules:

The roughest they method employs the prorty based apriori algorithm to mine association rules.The output of the feature selection continuous values such as  $A=8.5$ ,so the highest value is only applicable for applying this algorithm.

##### 4.2.1.STEP3: How to generate the rule for this algorithm:

In existing method first we check the A and B value and then combination of each features afterthat process goes on and then check the threshold value to produced the result of features.

In proposed that ,first to fix the threshold value and then check the combination between the features.Time complexity is less than existing method.

#### 4.3.STEP 4:Training set:

The output of the features of values in prorty based apriori algorithm will be stored in the training set.

#### 4.4.STEP 5:Testing set:

A standard approach to evaluate the accuracy of similarity values. we say that an image value descriptions matches a rule in the training set ,if the image feature satisfy the whole body of the rule. An image value descriptions partially matches a rule,it what kinds of stages will be produced.An image

value descriptions does not match a rule,if the image feature do not satisfy any part of the rules body.

## 5.CONCLUSION

This paper describes FUZZY based roughest theory is applied for reducing attributes and then increasing accuracy. It also proposed a priority based apriori for frequent item generation to reduce time complexity and improving accuracy. Then we apply classification for prediction of class labels. Our result shows , proposed work is better than existing one.

## 6. REFERENCES

[1] Li Q., Li F., Shiraishi J., Katsuragawa S., Sone S. And Doi K., \Investigation Of New Psychophys-Ical Measures For Evaluation Of Similar Images On Thoracic Computed Tomography For Distinctionbetween Benign And Malignant Nodules", Medical Physics,Vol. 30, No.30, Pp. 2584-2593

[2] Muller H., Michoux N., Bandond D. and Geissbuhler A., \A review of content-based image retrieval systems in medical applications-clinical bene-ts and future directions", Int J Med Informatics, vol. 73, pp. 1-23, 2004.

[3] Kawata Y., Niki N., Ohmatsu H., Kusumoto M., Kakinuma R., Yamada K., Mori K., Nishiyama H., Eguchi E., Kaneko M., and Moriyama N., "Pulmonary nodule classi-cation based on nodule retrieval from 3-D thoracic CT image database", Medical Image Computing and Computer-Assisted Intervention (MICCAI 2004).

[4] Lam M., Disney T., Raicu D. S., Furst J. and Channin D. S., \BRISC-An open source pulmonary nodule image retrieval framework", Journal of digital imaging, 2007.

[5].Lens MB, Dawes M. Global perspectives of contemporary epidemiological trends of cutaneous malignant melanoma. Br J Dermatol. 2004;150:179–85. doi: 10.1111/j.1365-2133.2004.05708.x. [PubMed] [Cross Ref]

[6].Schaffer JV, Rigel DS, Kopf AW, Bologna JL. Cutaneous melanoma: past, present, and future.J Am Acad Dermatol. 2004;51:S65–S69. doi: 10.1016/j.jaad.2004.01.030. [PubMed][Cross Ref]

[7] Fiona J. Gilbert, F.R.C.R., Susan M. Astley, Ph.D., Maureen G.C. Gillan, Ph.D., Olorunsola F. Agbaje, Ph.D.,Matthew G. Wallis, F.R.C.R., Jonathan James, F.R.C.R., Caroline R.M. Boggis, F.R.C.R., Stephen W. Duffy, M.Sc.,for the CADET II Group (2008). Single Reading with Computer-Aided Detection for Screening Mammography, The New

England Journal of Medicine, Volume 359:1675-1684 Full text ^ Effect of Computer-Aided Detection on Independent Double Reading of Paired Screen-Film and Full-Field Digital

[8].Screening Mammograms Per Skaane, Ashwini Kshirsagar, Sandra Stapleton, Kari Young and Ronald A. Castellino^ Taylor P, Champness J, Given-Wilson R, Johnston K, Potts H (2005). Impact of computer-aided detection prompts

[9].on the sensitivity and specificity of screening mammography. Health Technology Assessment 9(6), 1-70.^ Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C et al. Influence of computer-aided detection on

[10].performance of screening mammography. N Engl J Med 2007 April 5;356(14):1399-409. Full text^ Taylor P, Potts HWW (2008). Computer aids and human second reading as interventions in screening

[11].mammography: Two systematic reviews to compare effects on cancer detection and recall rate. European Journal of Cancer. doi:10.1016/j.ejca.2008.02.016 Full text ^ <http://www.cancer.org/downloads/CRI/6976.00.pdf>

[12]. Wu N, Gamsu G, Czum J, Held B, Thakur R, Nicola G: Detection of small pulmonary nodules using direct digital

[13].radiography and picture archiving and communication systems. J Thorac Imaging. 2006 Mar;21(1):27-31. PMID 16538152

[14] xLNA (x-Ray Lung Nodule Assessment)

[15] 10. ^ Petrick N, Haider M, Summers RM, Yeshwant SC, Brown L, Iuliano EM, Louie A, Choi JR, Pickhardt PJ. CT colonography with computer-aided detection as a second reader: observer performance study. Radiology 2008 Jan;246(1):148-56. Erratum in: Radiology. 2008 Aug;248(2):704. PMID 18096536

[16] ^ Halligan S, Altman DG, Mallett S, Taylor SA, Burling D, Roddie M, Honeyfield L, McQuillan J, Amin H, Dehmeshki J. Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection. Gastroenterology 2006 Dec;131(6):1690-9. Epub 2006 Oct 1. PMID 17087934

[17] R. Agrawal, T. Imielinski, and A. N. Swami, —Mining association rules between sets of items in large databases, in

Proc. 1993 ACM SIGMOD Int. Conf. Manage. Data – SIGMOD 93 SIGMOD 93, Washington, DC, 1993, pp. 207–216.

[18] M.X. Ribeiro, A.J.M. Traina, C.T. Jr., N.A. Rosa, P.M.A.

Marques, How to improve medical image diagnosis through association rules: The idea method, in: The 21th IEEE International Symposium on Computer-Based Medical Systems, Jyvaskyla, Finland, 2008, pp. 266–271.