HISTOGRAM NORMALIZATION TECHNIQUE FOR PREPROCESSING OF DIGITAL MAMMOGRAPHIC IMAGES

Adepoju, T. M.[1], *Adeyemo, T. T.[2] , Fagbola, T. M.[3], Omidiora, E. O.[4] and Olabiyisi, S. O.[5]

[2,4&5]Department of Computer Science and Engineering,
Ladoke Akintola University of Technology,
P.M.B. 4000, Ogbomoso, Oyo State, Nigeria.

[3]Department of Computer Science,
Federal University Oye-Ekiti, Ekiti State, Nigeria.

[1]Department of Computer Engineering Technology,
Federal Polytechnic Ede, Ede, Osun State, Nigeria.

[1]atemilola@gmail.com , [2]ttadeyemo@pgschool.lautech.edu.ng, [3]temitayo.fagbola@gmail.com
[4]eoomidiora@lautech.edu.ng [5]soolabiyisi@lautech.edu.ng

ABSTRACT

Digital mammogram has become the most efficient tool for early breast cancer detection modalities and pre-processing these images requires high computational capabilities. Pre-processing is one of the most important step in the mammogram analysis due to poor captured mammographic image qualities. Pre-processing is basically used to correct and adjust the mammogram image for further study and classification. Many image pre-processing techniques have been developed over the past decades to help radiologists in diagnosing breast cancer. Most studies conducted have proven that a pre-processed image is easier for radiologist to accurately detect breast cancer especially for dense breast. Different types of techniques are available for pre-processing of mammograms, which are used to improve image quality, remove noise, adjust contrast, enhance the image and preserve the edges within the image. This paper acquired 20 digital mammograms from Mammographic Image Analysis Society (MIAS) database and uses Histogram Normalization algorithm for pre-processing of the mammograms. A percentage of 95% was obtained. It was observed that the pre-processed mammographic images displayed breast abnormalities clearer with little or no noise.

Keywords— breast; cancer; mammogram; pre-processing; digital image; histogram

INTRODUCTION

Breast Cancer is one of the most common cancers, leading to cause of death among women, especially in developed countries (Kekre, Sarode, and Gharge, 2009). Clinical examination, mammography, ultrasound, and core biopsy are some common breast cancer detection methods (Medindia, 2006). Mammographic image is one of emerging technological advancements that have been used in diagnosing breast diseases or revealing cancer at early stages, where it is considered as the most effective method for the detection of early breast cancer (Karssemeijer, 1997). Mammographic

2013).

Digital mammography is the application of digital techniques on mammograms. Radiologists are depending on digital mammography for an alternative diagnostic method rather than using conventional screening programs because of its inherent problems. It can be said that an automated system will reduce problems with lesser number of false positive, false negative readings and increase the chance of early detection of abnormalities (Pisano and Shtern, 1994). With digital mammography, radiologists can adjust the brightness, enhance the contrast, and zoom in for closer view. Being able to manipulate medical images is one of the main benefits of digital technology (Antonie, Zaiane, and Coman, 2001).

Reading, interpreting and diagnosing the gray-scale mammographic image are difficult tasks, which require special training and experience. The main reason is that the breast tumors usually mix with the homogeneous tissues in the breast. The images provided by different patients have different dynamics of intensity, has high noise levels that can vary up to 10-15% of the maximum pixel intensity and present a poor contrast. Moreover the size of the significant details can be very small (calcifications) (Garge, and Bapat, 2009).

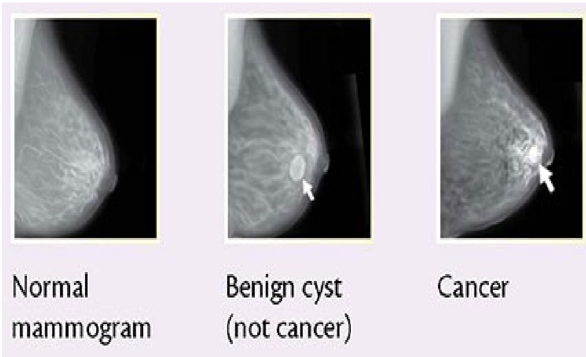

Figure 1: Different Mammographic images

Pre-processing techniques are necessary, in order to find the orientation of the mammogram, to remove the noise and to enhance the quality of the image. Before any image processing algorithm can be applied on mammograms, pre-processing steps are very important in order to limit the search for abnormalities without undue influence from background of the mammogram (Garge, and Bapat, 2009). Pre-processing techniques are used for reducing image noise, highlighting edges, contrast enhancement or removal of artifacts. Preprocessing basically indicates that the same tissue type may have a different scale of signal intensities for different images (Vendhan, Mathew, Brennan, Rajan, Kanimozhi, Mathews, Mathew and Boffetta, 2009). Preprocessing functions involve the operations that are normally required prior to the main data analysis and extraction of information, and are generally grouped as radiometric or geometric corrections (Singh and Al-Mansoori, 2000)

Several research works have tried to develop different enhancement and pre-processing techniques that could help the radiologists in the interpretation of the mammograms and could be useful for an accurate diagnosis. Most of these techniques undesirably enhance noise component in the image, and the lesions in mammographic images may appear quite subtle (Schiabel, Santos, and Angelo, 2008). Therefore, this paper presents a pre-processing method using Histogram Normalization (HN) that will improve the visual appearance of mammographic image. The main objective of this paper is to improve the quality of mammographic image, making it ready for further processing by removing the unwanted noise, improving the interoperability of the information present in the images and normalizing the image contrast.

This paper is organized into following sections. Section 2 presents literature review on pre-processing techniques for enhancing of mammographic images and other area of pattern recognition. Section 3 discusses the methodology and algorithm used for the pre-processing. Results and discussion is presented in section 4 and concluded in section 5.

DIGITAL IMAGE

A digital image is a discrete space composed of small surface elements called pixel. Each one of these elements contains a value or a set of value coding the intensity level at each position. A digital image can be acquired with a great number of different devices such as a camera, an MRI machine or any kind of device with a sensor able to capture light intensity (Chanda, and Majumder, 2004). Because of its discrete nature, the theory used to process digital image will rely on discrete domain, even if the analogy with the continuous domain is possible.

A gray scale image is a digital image in which each pixel only contains one scalar value which is its intensity. The number of possible levels (intensity values) depends on the numerical type encoding the image. For example, an image encoded with n=8 bits will only have $L = 2^8 = 256$ possible intensity values going from 0 representing black to $L-1 = 255$ representing white (Jensen, 1996).

Digital Mammograms are medical images requires a preparation phase to improve the quality of the image. The main objective during this process is in preparing the image and makes it ready for further processing by removing the irrelevant, artifacts and unwanted parts in the background of the mammogram as displayed in Figure 2 (Sastre, 2011).

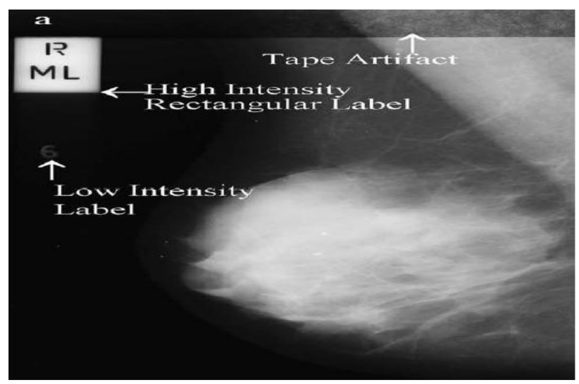

Figure 2: Type of noises present in mammogram

Histogram

The histogram of an image normally refers to a histogram of the pixel intensity values. It provides information about intensity distribution of pixels in the image. Histogram is a graph showing the number of pixels in an image at each different pixel intensity value found in that image as shown in Figure 3 (Raba, Oliver, Martí, Peracaula, and Espunya, 2005).

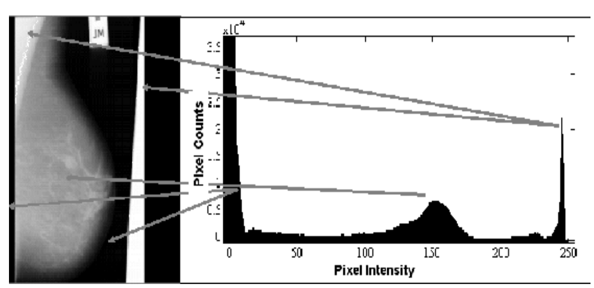

Figure 3: Mammogram and its histogram showing the number of pixel intensity

For an 8-bit grayscale image there are 256 different possible intensities, and so the histogram will graphically display 256 numbers showing the distribution of pixels amongst those grayscale values. The histogram of a digital image $h(r_i)$ is expressed as:

$$h(r_i) = n_i \qquad for\ i = 0,1,....,L-1 \qquad (1)$$

Where r_i is the i^{th} gray-level in the image for a total of L gray values and n_i is the number of occurrences of gray-level r_i in the image. If a histogram is expressed in terms of the probability of occurrence of gray-levels, it can be expressed as:

$$p(r_i) = \frac{n_i}{n} \qquad (2)$$

Where n is the total number of pixels.

Contrast generally refers to the difference in luminance or grey level values in an image. It can be defined as the ratio of the maximum intensity to the minimum intensity over an image. Contrast ratio has a strong bearing on the resolving power and detectability of an image. Thus, a histogram is a plot of $h(r_i)$ or $p(r_i)$ versus r_i showing the image contrast as displayed in Figure 4 (Sarage and Jambhorkar, 2011).

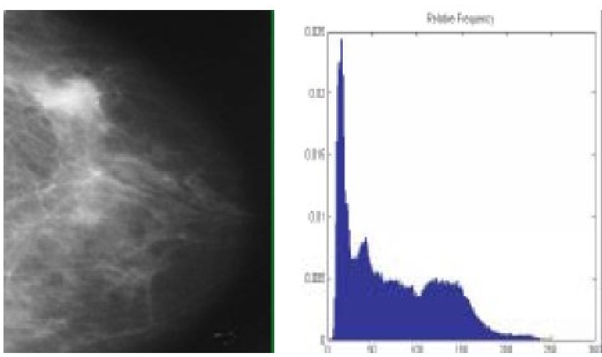

Figure 4: Cropped Mammogram and its gray level histogram

RELATED WORKS

In digital mammogram pre-processing, the images are mostly affected by various noises. The function of the pre-processing technique is to improving the appearance of an image without referring with the conditions of image degradation process (Sakellaropoulos, Costaridou, and Panayiotakis, 2003). This will improve medical analysis, in order to acquire a good quality images so that radiologists can make use of these images to arrive at an accurate conclusions.

Image pre-processing researchers have developed various algorithms, Rahmati, Hamarneh, Nussbaum and Adler (2010) used fuzzy logic to develop the CLAHE algorithm. Fuzzy Contrast Limited Adaptive Histogram Equalization (FCLAHE) was used to enhance contrast mammographic image. The filter (used in preprocessing stage) eliminates noise and intensity in homogeneities in the background while retaining the natural gray level variations of mammographic images within suspicious lesions, before Segmentation. Sara and Mashallah (2011) represented pre-processing of Mammogram images using Grey level thresholding and gradient based methods.

Elsawy, Sayed, Farag, and Gouhar (2012) presented the contrast enhancement algorithm for mammographic image. The algorithm performs Band-Limited Histogram Equalization (BLHE) for certain intensity band of the mammographic image histogram. Mohan and Mahesh (2013) proposed a PSO based local contrast modification CLAHE algorithm. The CLAHE method can limit the noise enhancement. Local contrast modification CLAHE algorithm is optimizes using the PSO algorithm. The new histogram is different from the normal histogram because intensity of each pixel is limited by user selectable maximum.

Hussain (2014) proposed the method that used Fast Dyadic Wavelet Transform (FDyWT) for the enhancement of mammographic image for multi-scales analysis, normalized Tsallis entropy for reduction the noise , a non-linear function for contrast enhancement and a power law transformation for background suppression. Adepoju, Ojo, Omidiora, Olabiyisi, and Bello (2015) pre-processed digital mammograms by removal of artifacts using region description technique.

Pre-processing is also applied in other field of pattern recognition. In textured textile images, Anitha and Radha (2010) described how different preprocessing techniques like contrast adjustment, intensity adjustment, histogram equalization, binarization and morphological operation are applied. In face recognition and verification, Leszczynski (2010) analyzed 14 normalization algorithms based on histogram normalization and illumination properties of the digital images. In Classification of Textures Based on Features Extracted from Preprocessing Images on Random Windows, Reddy, Suresh, Mani and Kumar (2009) described the image preprocessing methods applied on sequential window and random window.

Gupta, Jacobson and Garcia (2007) devised the binarization, pre-filtering, and post-binarization, de-noising for the optical character recognition. In brain tumor detection Wang, Liu, Xiang and Yang (2006) presented a rough set feature selection and rule induction for prediction of malignancy degree in brain glioma using statistical structure analysis method for brain Images. In Preprocessing of a fingerprint image captured with a mobile camera, Lee, Lee, Kim and Kim (2006) described the Tenengrad based method using the Sobel operator for pre-processing of fingerprint image.

METHODOLOGY

Database

For the experiment, digital mammographic images were acquired from MIAS (Mammographic Image Analysis Society) database. MIAS images were originated as the product of the film-screen mammogram process in the United Kingdom National Breast Screening Program. The films have been digitized using a Joyce-Lobel scanning, a device linear in the optical density range 0-3.2. The image size the mammograms acquired from MIAS is 1048 x 1048 pixels.

Histogram Normalization (HN)

Normalizing a histogram is a technique involving transformation of the discrete distribution of intensities into a discrete distribution of probabilities. To do so, each value of the histogram is divided by the number of pixel. HN transforms an n-dimensional grayscale image $I : \{X \subseteq Rn\} \rightarrow \{\min, \ldots, \max\}$ into a new image $I_N : \{X \subseteq Rn\} \rightarrow \{newmin, \ldots, newmax\}$ with intensity values in the range {newmin, newmax}. The linear normalization of a grayscale digital image is performed by

$$I_N = (I - \min)\frac{new\max - new\min}{\max - \min} + new\min \quad (3)$$

Pixel intensity numbers are scaled such that either the max/min will be equal to a specific number (usually 1) in order to normalize the contrast of the image. In this research, the histogram of the original image is transformed by using its normalized cumulative sum, if the overall brightness of an image is controlled by a level l and the range is controlled by a gain *k*, the brightness of the points in the new image *N*, can be related to the brightness in old image *O* by using

$$N_{x,y} = K \times O_{x,y} + 1 \quad \forall x,\ y \in 1,\ N \quad (4)$$

Then the intensity values of the original image are mapped to new intensity to give a uniform histogram of intensity values and for normalizing the intensity using

$$N_{x,y} = \frac{N_{\max} - N_{\min}}{O_{\max} - O_{\min}} \times (O_{\max} - O_{\min}) + N_{\min} \quad \forall x, y \in 1,\ N \quad (5)$$

The idea of normalizing a histogram is to stretch and/or redistribute the original histogram using the entire range of discrete levels of the image, in a way that an enhancement of image contrast is achieved. The reference image histogram is stretched and shifted in order to cover all the grayscale levels in the input image. In this research, the pre-processing of the mammograms involves image reduction, masking, cropping and enhancement as demonstrated in Figure 5.

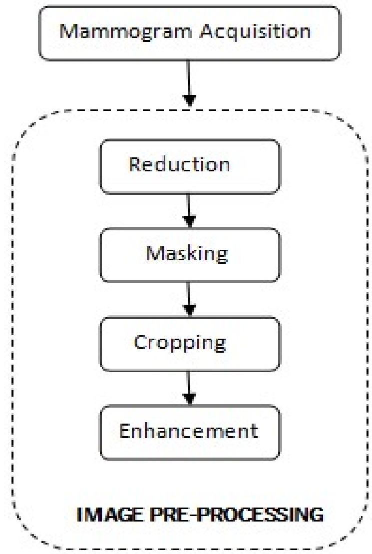

Figure 5: Block diagram of the proposed method

Reduction: Reduction of the original image size was done resulting in an image of pixel size of 512 X 512. The objective of the reduction is to reduce the preprocessing time which will be too high, if the original image is used.

Masking: This involves the estimation of the breast boundary using adaptive thresholding. This was used to compute the outline of the breast region and generate the mask.

Cropping: Mammographic images were cropped such that the breast region is aligned to its respective side. The images were cropped from the edges to remove all unwanted element.

Enhancement: Histogram Normalization (HN) was used to enhance the image by normalizing the contrast. HN was used to enhance the image contrast that scales the gray level. This improves the visualization effect of the original image, making the Region of Interest (ROI) apparent.

RESULTS AND DISCUSSION

Mammographic image 'mdb001' – 'mdb020' were selected from the publicly acquired database (MIAS). Figure 6(a) and (b) displayed 'mdb005' before and after preprocessing. Out of 20 mammograms used for this experiment. 19 of them were properly preprocessed. Due to variation of gray levels, breast tissue or the contrast of intensities in the image intensity, 'mdb018' displayed a poor contrast image after pre-processing. The image appeared faint as displayed in Figure 7.

Figure 6(a). Acquired mammogram (mdb001) with noise and artifacts

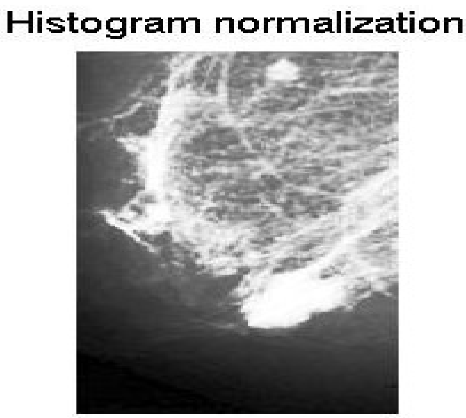

Figure: 6(b). Preprocessed mammogram (mdb001)

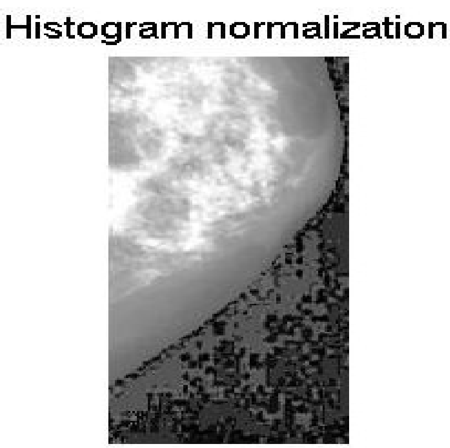

Figure 7: Preprocessed mammogram (mdb018)

The percentage of the experiment is 95%

$$\frac{19}{20}\times 100 = 95\ \%$$

It was observed that this research can play a very important role in improving the results of image analysis including segmentation, feature extraction and classification. Region of interest (ROI) can be easily identified from the preprocessed image. Normal and abnormal mammograms can be assumed with the pre-processed image because the ROI is brighter and clearer.

CONCLUSIONS

Pre-processing techniques are used for enhancing the content of medical image based on removal of special markings and noise. Removal of special markings and noise existing in medical images will increase the quality of image segmentation and accurate classification. The techniques are useful in modifying the gray level values of individual pixels to better visual quality of the entire image. This will help the radiologist and surgeons by improving the possibility of interpretation and perception of mammographic image information. It also assists to detect irregular shaped mammograms. Mammography breast cancer images have the ability to assist physician in detecting disease caused by cells normal growth. Developing algorithms and techniques to analyze these medical images may also assist physicians in their daily work. This paper shows that the outcome of applying Histogram Normalization processing technique on digital mammograms.

REFERENCES

Adepoju, T. M., Ojo, J. A, Omidiora, E. O., Olabiyisi, S. O. and Bello, T. O. (2015). “Detection of tumour based on breast tissue categorization”, British Journal of Applied Science & Technology, vol. 11, pp.1-12.

Anitha, S. and Radha, V. (2010). “Comparison of image preprocessing techniques for textile texture images. International Journal of Engineering Science and Technology, vol. 2, pp. 7619-7625.

Antonie, M. L., Zaiane, O. R and, Coman, A. (2001). “Application of data mining techniques for medical image classification”, MDM, pp. 94-101.

Chanda, B. Majumder, D. D. (2004). “Digital image processing and analysis”, PHI Learning Pvt. Ltd., pp. 123-129.

Dehghani, S. and Dezfooli, M. A. (2011). “ A method for improve preprocessing images mammography” International Journal of Information and Education Technology, vol. 1, pp. 90-96.

Elsawy, N. Sayed, M. S Farag, F. and Gouhar, G. K. (2012). “Band-limited Histogram Equalization for Mammograms Contrast Enhancement”, IEEE Cairo International Conference in Biomedical Engineering (CIBEC), pp. 154-157.

Garge, D. M. and Bapat, V. N. (2009) “A low cost wavelet based mammogram image processing for early detection of breast cancer”, Indian Journal of Science and Technology, vol. 1, pp. 63-67.

Gajalakshmi Vendhan, Aleyamma Mathew, Paul Brennan, Balakrishnan Rajan, Vendhan C. Kanimozhi, Anitha Mathews, Beela S. Mathew and Paolo Boffetta (2009). “Breastfeeding and breast cancer risk in India: a multicenter case-control study”, International Journal on Cancer, vol. 125, pp. 662-665.

Gupta, M. R. Jacobson, N. P. and Garcia, E. K. (2007). “OCR binarization and image pre-processing for searching historical documents”, Pattern Recognition, vol. 40, pp. 389-97.

Hussain, M. (2014). “Mammogram Enhancement Using Lifting Dyadic Wavelet Transform and Normalized Tsallis Entropy”, Journal of Computer Science and Technology, Vol. 29, pp. 1048–105.

Jensen, J. R (1996). “Introduction to Digital Image Processing”, A Remote Sensing Perspective, Practice Hall, New Jersey.

Karssemeijer, N (1997). “Computer-Assisted Reading Mammograms”, European Radiol., vol. 7, pp. 743–748.

Kekre, H. B., Sarode, T. K. and Gharge, S. M. (2009). "Tumor detection in mammography images using vector quantization technique", International Journal of Intelligent Information Technology Application, vol. 2, pp. 237-242.

Lee, C., Lee, S. Kim, J. Kim, S. J. (2006). "Preprocessing of a fingerprint image captured with a mobile camera", Advances in Biometrics, [Springer Berlin Heidelberg], pp. 348-355.

Leszczyński, M. (2010). "Image preprocessing for illumination invariant face verification", Journal of Telecommunications and Information Technology, vol. 1, pp.19-25.

Medindia (2006) "Breast cancer in India rising rapidly". http://www.medindia.net/news/view_news_main

Mohan, S. and Mahesh, T. R. (2013). "Particle Swarm Optimization Based Contrast Limited Enhancement for Mammogram Images", In Intelligent Systems and Control (ISCO), [7th International Conference on IEEE], pp. 384-388.

Pisano, E. and Shtern, F. (1994). "Image processing and computer-aided diagnosis in digitalmmammography," in State of the Art of Digital Mammographic Image Analysis, Singapore: World Scientific, vol. 7,, pp. 280–291.

Raba, D., Oliver, A., Martí, J., Peracaula, M. Espunya, J. (2005). "Breast segmentation with pectoral muscle suppression on digital mammograms", Pattern Recognition and Image Analysis, pp. 471-478, Springer Berlin Heidelberg.

Rahmati, P. Hamarneh, G. Nussbaum, D. and Adler, A. (2010). "A new preprocessing filter for digital mammograms", InImage and Signal Processing, [Springer Berlin Heidelberg], pp. 585-592.

Reddy, B. R., Suresh, A., Mani, M. R. and Kumar, V. V. (2009). "Classification of textures based on features extracted from preprocessing images on random windows", International Journal of Advanced Science and Technology, vol. 9 pp. 123-130.

Sakellaropoulos, P. Costaridou, L. and Panayiotakis, G. (2003). "Wavelet-Based Spatially Adaptive Method for Mammographic Contrast Enhancement", Physics in Medicine and Biology, Vol. 48, pp. 787-790.

Sarage, G. N. , Jambhorkar, S. (2011). "Enhancement of Mammography Images for Breast Cancer Detection using Histogram Processing Techniques" international journal of computer science and technology, vol. 2, pp. 177-179

Sastre, T. J. (2011). "Segmentation of the breast region with pectoral muscle suppression and automatic breast density classification", M.sc thesis computer science, université Catholique de Louvain, pp. 34-37.

Schiabel, H., Santos, V. T and Angelo, M. F. (2008). "Segmentation technique for detecting suspect masses in dense breast digitized images as a tool for mammography CAD schemes",[In Proceedings of the 2008 ACM symposium on Applied computing], pp.. 1333-1337.

Singh, S. and Al-Mansoori, R.(200). "Identification of regions of interest in digital mammograms", Journal of Intelligent Systems, Vol.10, pp. 183-217.

Tarver, T. (2013). "Cancer facts & figures", American Cancer Society (ACS). Journal of Consumer Health on the Internet, PP.366-367.

Wang, X. Yang, J. Jensen, R. and Liu, X. (2006). "Rough set feature selection and rule induction for prediction of malignancy degree in brain glioma", Computer methods and programs in biomedicine, Vol. 3, pp. 147-156.