Statistical Analysis of the Difference in the Centrosomal Features of Untreated Breast Cancer Cells and Normal Cells

Sushma S[1], Latha KC[2], Balasubramanian S[3], Sridhar R[4]

Abstract The Bosom disease is the most widely recognized type of tumour in people. The most established recognizable proof and meaning of tumour were recorded in Egypt in around 1600 BC. From that point forward this illness has been explored and examined to stay away from results brought about by it yet at the same time this ailment is considered as a standout amongst the deadliest infections of all times, as mortality rate associated with the disease reached 40,000 in 2012 in the US. In the advanced therapeutic science, there are a lot of recently formulated procedures and strategies for the convenient recognition of breast malignancy. Most of these procedures utilize exceptionally propelled innovations, for example, medicinal image handling. This exploration study is an endeavour to highlight the strategies for identification of breast malignancy in the light of imaging techniques and gives an outline of the reasonableness, unwavering quality, and results of every method. This paper is used to distinguish untreated breast cancer cells from normal cells through the statistical analysis of centrosomal features extracted from cell images.

Keywords Breast Cancer . Centrosome image . mammogram . K-Nearest Neighbour Algorithm . Statistical Analysis.

1 Introduction

Breast Cancer is considered as the second most common disease in the world, caused by the development of tumour cells. The most recent research from National Tumor Organization (NTO) reveals that around 232,340 females and 2,240 males suffered from breast cancer. Malignant tumour in breast starts growing and enters into the tissues. This disease generally occurs in females however men can also get affected from it. Breast cancer is most common among females of the UK with approximately 54 thousand cases diagnosed in 2014. However, during 2012-2014, breast cancer was diagnosed mostly in aged people i.e. above 60 years of age. Many researchers worked on this topic and provided important details about breast cancer. Existing information is very helpful for early diagnosis of breast cancer and also for those researchers who are willing to do further studies related to this topic. The details demonstrated that this disease is extremely normal in ladies and despite a lot of work being done, there is still a lot more that needs attention. Medicinal examination focusing on breast growth is not new and it roots back to the sixteenth century. Communication and progression need a lot to make improvements in the medicinal field because the incidence of this disease has continued and is still thought to be one of the most severe infections. Medical Image Processing (MIP) is the latest technique associated with information technology, used for the diagnostic purpose in the retrospective field. This process is not just restricted to malignancy illness, rather it also helped incredibly in the determination of various types of illnesses. This latest technology MIP shows result in measurements, with the assistance of image processing procedures. Mammograms differentiate indicators emitted from anomalous and typical tissues, providing images of tumours. Early detection can help in legitimate finding and treatment for minimizing the danger of most undesirable results of this disease (death). Images processed by medical equipment are distinctive in nature and need special supervision before arriving at a specific conclusion. Likewise, there are many different sources (machines) which create the pictures of human body parts. An uncommon strategy to manage unique kind of pictures is required for prompting diverse classes and mechanisms for diagnosis. This study was performed to determine various forms of breast tumour imaging techniques and their effects. This research work attempted to give the performance, precision and moderateness level of the discussed techniques. Each discussed strategy is interesting in its tendency and targets an uncommon type of situation. This paper also deals with the advantages and disadvantages of these techniques (Nagaraj, Paga, and Lamichhane 2014).

The images are classified on the score of particular aspects, the content-based image retrieval systems (CBIR) (Wan et al. 2014) is very useful and efficient to perform scoring task. For example, in a great database, the images can be divided into different classes as follows: landscapes, buildings, animals, faces, artificial images, etc. Many colour based image classification methods use colour histograms. In (Lin and Shyu 2012). feature vectors are generated using the Haar wavelet and Daubechies wavelet of colour histograms. Another histogram based approach can be found (Wang 2012), where the blob world is used to search similar images. Some methods include fuzzy features (Babu, Venkateswarlu and Chintha 2014; Thirumuruganathan 2010), invariant moment's features (Thirumuruganathan 2010), and structural and statistical features (Kumar, Srivastava and Srivastava 2015; Thirumuruganathan 2010) extraction. Aim of this paper is to evaluate the histogram based classification approach, which is efficient, quick and robust enough. This study takes part in the classification of images by using colour features of histogram. The advantage of this approach is to make a comparison of histogram features with other known methods. Histogram appears to be much faster and more efficient than other used methods.

2 Theoretical background

In this section, the theoretical background is summarized according to classification method. The certain subsections are based on Haidekker (Haidekker 2011).

2.1 Centrosome

The centrosome is an organelle that serve as the primary region for arrangement of microtubules. The centrosome copies itself just once in the middle of every cycle with duplication starting close to the G1-S and finishing at the G2 stage. Copied centrosomes are further divided into two mitotic shafts that regulates the whole mechanism of mitosis. The hereditary information of the human cells is kept by centrosomes, guaranteeing legitimate segregation of chromosomes. Human diseases can also be caused by the occurrence of different abnormalities in centrosomes for instance, malignancies of the lung, bosom, gallbladder, bone, pancreas, colon, rectum, head, neck, prostate, and ovaries. Recent information demonstrates the major reason behind different human cancers to be a loss of centrosomal functioning, also responsible for genetic instability.

Aneuploidy of non-small cell breast tumour is connected with centrosomal abnormalities. In breast cancer, critical discoveries proposed that centrosomal variations from the normal may occur at a moderately early phase. Also, it was demonstrated that stepwise movement of centrosome defects is connected with adjacent tumour. This tumour shows progression towards more propelled stage and also accelerates metastatic procedure of lung carcinoma cells. This article took a view of breast cancer and accommodates through, an objective, quantitative evaluation of centrosomal abnormalities. This quantitative centrosomal evaluation shows that growth of cancer cells which remains untreated can be effectively recognized by a different methodology. The methodology includes identification of tumour cells for cancer treatment, early findings and visualization by using different medical equipment (Zulkepli et al. 2012).

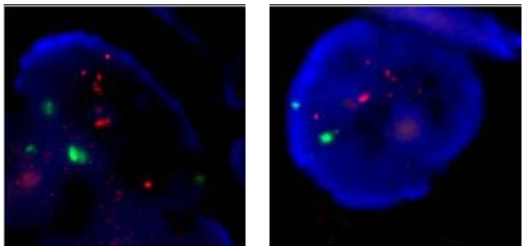

Fig 1. Centrosome image of mammogram (normal and benign)

2.2 Image Acquisition

Centrosomal images were acquired with the help of Analytic Microscopy Core (AMC) at the H. Lee Moffitt Cancer Center. Lungs cancer cells A549 and normal bronchial epithelial cell BEAS 2B were grown in RPMI (Invitrogen, Carlsbad, California, U.S.A.) with 10% foetal bovine serum and bronchial epithelial growth medium (BEGM) (Lonza Walk- ersville Inc., Walkersville, Maryland, U.S.A.) supplemented with a BEGM bullet kit, respectively. Cells were plated and grown on coverslips in a 6- well plate at 37°C with 5% CO2. Cells were fixed using 4% paraformaldehyde solution for 30 minutes at 4°C and permeabilized using 0.5% Triton X solution (Sigma-Aldrich, St. Louis, Missouri, U.S.A.). Following blocking with 2% bovine serum albumin, cells were stained with γ-Tubulin antibody (Sigma-Aldrich). Cells were then incubated with Alexa Fluor 594 secondary antibody (Invitro- gen) and mounted using Prolong Antifade with DAPI (Invitrogen). A DMI6000 inverted Leica TCS AOBS SP5 tandem scanning confocal microscope (Leica Microsystems GmbH, Wetzlar, Germany) was used to image the cells, under a 100× oil immersion objective with scanning speed of 100 Hz per each 2048 × 2048 frame (Figure 1a and b). The LAS AF software suite (Leica Microsystems GmbH) was used to image the cells and compile the maximum projections from Z-stacks. The acquired image has a resolution of 75.7 nm (Zulkepli et al. 2012).

2.3 Histogram Features

The histogram features read the frequencies of the pixels which can be helpful for detecting abnormalities in cells. The histogram plots a picture of the dark level qualities and has the power to estimate shading channel. Statistical analyses of cancer cells is predicted by the help of gray-level values. The shape of the histogram provides data about the way of the picture, or sub-image on the likelihood that is rational about an item inside the picture. For instance, an exceptionally contract histogram infers a low differentiation picture, a histogram skewed towards the top of the line infers a splendid picture, and a histogram with two noteworthy crests, called bimodal, infers an item that is conversely with the foundation (Zulkepli et al. 2012).

The histogram will be considered as a measurable component, where the histogram is utilized as a model of the likelihood circulation of the power levels. These measurable components give a data about the qualities of the force level circulation for the picture. Characterizing the principal request histogram likelihood, P (g), as:

P (g) =N (g)/M

M is the quantity of pixels in the picture (if the whole picture is under thought then M = N 2 for N picture), and N (g) is quantity of pixels at dark level g. Likewise, with any dissemination, every one of the qualities for P (g) are not exactly or equivalent to 1, and the total of all the P (g) qualities is equivalent to 1. The components in view of the primary request histogram likelihood are the mean, standard deviation, skew, vitality, and entropy.

The mean is the normal quality, so it tells something about the general brilliance of the picture. A splendid picture will have a high mean, and a dull picture will have a low mean. L will be utilized as the aggregate number of force levels accessible, so the dark levels range from 0 to L 1. For instance, for ordinary 8-bit picture information, L is 256 and ranges from 0 to 255.

The skew measures the asymmetry about the mean in the intensity level distribution. It is defined as:

L-1g = gP (g) = I(r, c)

g=0

If the tail of histogram spreads to the right side than the skew will be positive. The skew will be considered negative if the tail of the histogram spreads to the left (negative). Skew can also be measured by calculating mean, mode and standard deviation as an alternative method where, the mode is defined as the peak, or highest, value:

SKEW = g mode/g

This method is more computationally efficient, especially considering that, typically, the mean and standard deviation have already been calculated.

The energy measure tells about the disturbance of intensity level:

L-1

ENERGY

P (g) g=0

The energy measure has a maximum value of 1 for an image with a constant value and gets increasingly smaller as the pixel values are distributed across more intensity level values (remember all the P (g) values are less than or equal to 1). The larger this value is, the easier it is to compress the image data. If the energy is high, it informs that the number of intensity levels in the image is limited, that is, the distribution is concentrated in only a small number of different intensity levels.

The entropy is a measure that tells about the number of bits needed to code the image data and is given by ([8], 2014, 31):

L-1 ENTROPY=P (g) log2 [P (g)]

g=0

As the pixel values in the image are distributed among more intensity levels, the entropy increases. Image shown below has higher entropy as compared to simple image. This measurement tends to vary inversely with the energy (Zulkepli et al. 2012).

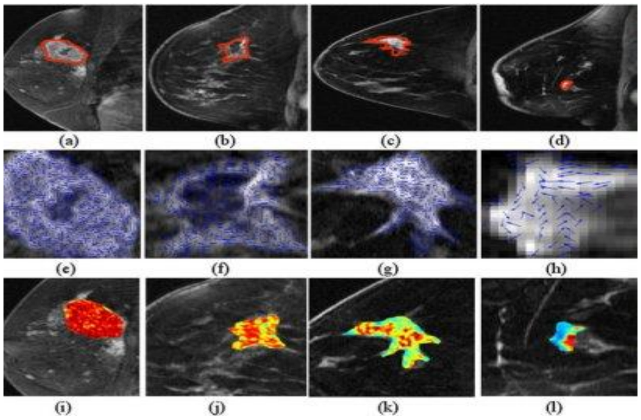

Fig 2. A complex image is shown higher entropy than a simple image.

2.4 Feature vector and Feature Space

A component vector is one strategy to speak to a picture by discovering estimations on an arrangement of elements. A vector specified as n-dimensional contains these estimations, where n showed the quantity of elements. The estimations might be typical, numerical, of both. A case of a typical element is shaded as "blue" or "red"; a case of a numerical component is the territory of an item.

In the event, by taking a typical element and appoint a number to it, it turns into a numerical component. Care must be taken is allotting numbers to typical components, so that the numbers are allocated genuinely. For instance with shading must think about the tint by its name, for example, "orange" or "maroon". For this situation, technician could play out a HSL change on the RGB information, and utilize the H (shade) esteem as a numerical shading highlight. In any case, with the HSL change the tint esteem ranges from 0 to 360 degrees, and 0 is "by" 360, so it is invalid to think about two hues by just subtracting the two shade values.

The component vector can be utilized to characterize an article or give us consolidated more elevated amount picture data. Connected with the component vector is a numerical abdominal muscle traction called an element space, which is additionally n-dimensional and is made to permit representation of highlight vectors, and connections between them. With two-and three-dimensional component vectors, it is displayed as a geometric develop with opposite tomahawks and made by plotting every element measurement along one pivot. For n-dimensional component vectors it is a conceptual scientific development called a hyperspace. It should be seen that the production of component space permits us to characterize separation and closeness measures which are utilized to analyse highlight vectors and help in the characterization of obscure examples.

2.5 Classification Algorithms and Methods

The simplest algorithm is used for identifying a sample taken from the test set known as the Nearest neighbour method. The objective of this study is to take a view of each and every sample run in the training set. It can be done by using a distance measure, a similarity measure, or a combination of measures. After taking a comparative view, the unknown object is then identified as belonging to the same class and also detected as the closest sample in the training set. If a distance measure is used, the unknown object indicated by smallest number and if similarity measure is used, it showed largest number. This process is computationally intensive and not very robust (Zulkepli, Eldabi, and Mustafee 2012).

This study can make the Nearest Neighbour method more strong by selecting a group of close feature vectors rather than taking a closest sample present in the training set. This is called the K-Nearest Neighbor method, for example, K = 5. The class that occurs most often in the set of K-Neighbours get assigned the unknown feature vector. This is still very computationally intensive. Make it possible that each and every unknown sample must compared in the training set to get maximum success. This study can reduces the computational burden by using Nearest neighbor method. Here, it is found that the sample of centroids from the training set were given to each class and then compare it to the unknown samples with the given centroids only. The centroids present in training set were calculated by finding the average value of it (Zulkepli et al. 2012).

3 Methods

The main purpose of this study is to find out centrosomal features by comparing untreated cancer cell with a normal cell. The Study design used for this research is statistical analyses by using histogram features.

4 Study design

The advanced research indicates that human cancer cell development is based on centrosomal abnormalities. For the analysis of centrosome, high-resolution images were acquired which clearly specify normal and untreated cancer breast cells. After the centrosomal images were segmented, statistical features were identified.

5 Experiments

During the experiment, 10 training images were used. These images were divided into two size classes: namely Cancer cell image of centrosome and normal cell image of the centrosome.
From each image classes, Statistical texture features are calculated. Further, for each training set, the histograms of the three colour channels were generated and the above-mentioned histogram features were calculated.

The obtained dimensional feature vectors are classified using K-Nearest Neighbor Classifier. Further the Accuracy is measured for different values of K. MATLAB software was used for the coding of algorithms because this system works computationally very fast as compared to others and the code generation is very simple in it.

6 Results

Figure 3 indicates the Statistical Texture features (Namely Mean, Energy, Entropy, etc...) for the Cancerous Centrosome Images.

X axis Indicates the 18 Features for the 3 different planes of Colour image Y axis Indicates the Values for each Features.

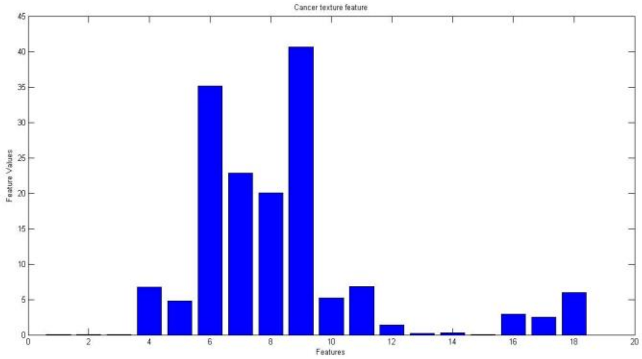

Fig. 3. Statistical Texture features for Cancerous Centrosome

Figure 4 indicates the Statistical Texture features (Namely Mean, Energy Entropy, etc...) for the Normal Centrosome Images.

X axis Indicates the 18 Features for the 3 different planes of Colour image Y axis Indicates the Values for each Features.

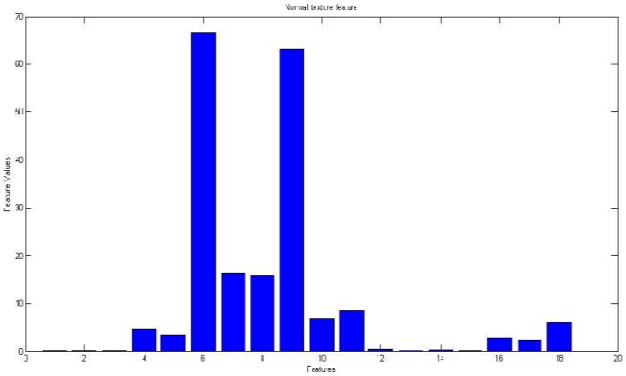

Fig. 4. Statistical Texture features for Normal Centrosome

Figure 5 indicates the Recognition Performance for different Values of K (Since using K-NN Algorithm for Classification).

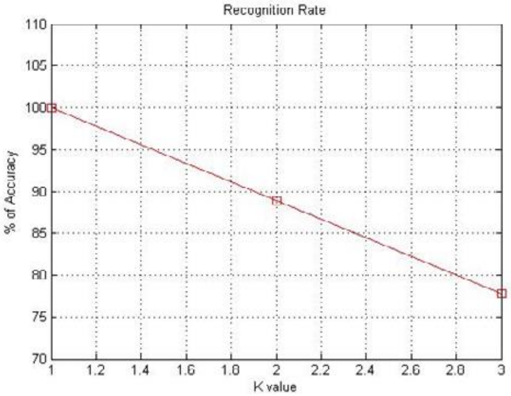

Fig. 5. Recognition Performance for different values of K

Centrosomal features of statistical texture namely which mean entropy of both cancer and normal centrosomal images indicates the recognition performance for different values of K (K-NN algorithm for classification) (Nagaraj, Paga, And Lamichhane 2014). It shows that “As the K value Increases the Accuracy Decreases”.

7 Discussion

This case discusses statistical analyses that were constructed to predict the breast cancer case-control status by using histogram features (Wolff et al. 2013). Histogram features are commonly known as image analysis that was automatically calculated from the affected areas of the breast. Textural structures can further assists at various levels. After finding the statistical features (Delen 2009), the image was further computed on three different scales. One of them is original mammograms and rest of the two are reduced version in which pixel size of the image is 0.5cm-1.0cm per pixel (Nagaraj, Paga and Lamichhane 2014). It takes only 12 seconds to compute all three image scales which can explain the best mammographic image for detecting cancer cells (Pettersson et al. 2014; Bleyer and Welch 2012) in the affected areas of the breast (Buciu and Gacsadi 2011). The statistical analysis between normal cells and untreated cancer centrosomes shows the significant difference. The work presented in this manuscript is still in progress to identify more procedures distinguishing malignant from normal lung cells and also use some advanced staining technique for finding centrosomal abnormalities. These centrosomal abnormalities can cause chromosomal instability (Godinho and Pellman 2014), because development of this objective rather than tumorigenesis may facilitate early diagnosis, improved prognosis and early detection.

8 Conclusion

In this paper an approach of the statistical feature of centrosome comparison of untreated and normal cells using a nearest neighbour algorithm. Histogram features generated computationally.

It has been statistically proved that the k value increases the accuracy decreases.

Abbreviations NTO- National Tumor Organization; MIP-Medical Image Processing; CBIR-Content-based Image Retrieval; AMC- Analytical Microscopy Core; BEGM- Bronchial Epithelial Growth Medium

Acknowledgements The authors thank all of the analysts who participated in this research and those who make innovation trends in computing and communication.

Competing interests The author states that there is no competing interests regarding this research.

References

Babu, U.R., Venkateswarlu, Y. and Chintha, A.K., 2014, February. Handwritten digit recognition using K-nearest neighbour classifier. In *Computing and Communication Technologies (WCCCT), 2014 World Congress on* (pp. 60-65). IEEE..

Bleyer, A. and Welch, H.G., 2012. Effect of three decades of screening mammography on breast-cancer incidence. *New England Journal of Medicine*, *367*(21), pp.1998-2005.

Buciu, I. and Gacsadi, A., 2011. Directional features for automatic tumor classification of mammogram images. *Biomedical Signal Processing and Control*, *6*(4), pp.370-378.

Delen, D., 2009. Analysis of cancer data: a data mining approach. *Expert Systems*, *26*(1), pp.100-112.

Godinho, S.A. and Pellman, D., 2014. Causes and consequences of centrosome abnormalities in cancer. *Phil. Trans. R. Soc. B*, *369*(1650), p.20130467.

Haidekker, M., 2011. *Advanced biomedical image analysis*. John Wiley & Sons.

Kumar, R., Srivastava, R. and Srivastava, S., 2015. Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features. *Journal of medical engineering*, *2015*.

Lin, L. and Shyu, M.L., 2012. Weighted association rule mining for video semantic detection. *Methods and Innovations for Multimedia Database Content Management*, p.12.

Nagaraj, H., Paga, P. and Lamichhane, K., 2014. Early breast cancer detection using statistical parameters. *Int J Res Engineer Technolo*, *2*, pp.31-6.

Pettersson, A., Graff, R.E., Ursin, G., dos Santos Silva, I., McCormack, V., Baglietto, L., Vachon, C., Bakker, M.F., Giles, G.G., Chia, K.S. and Czene, K., 2014. Mammographic density phenotypes and risk of breast cancer: a meta-analysis. *Journal of the National Cancer Institute*, p.dju078.

Thirumuruganathan, S., 2010. A detailed introduction to K-nearest neighbor (KNN) algorithm. *Retrieved March, 20*, p.2012.

Wan, J., Wang, D., Hoi, S.C.H., Wu, P., Zhu, J., Zhang, Y. and Li, J., 2014, November. Deep learning for content-based image retrieval: A comprehensive study. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 157-166). ACM..

Wang, J.Z., 2012. *Integrated region-based image retrieval* (Vol. 11). Springer Science & Business Media.

Wolff, A.C., Hammond, M.E.H., Hicks, D.G., Dowsett, M., McShane, L.M., Allison, K.H., Allred, D.C., Bartlett, J.M., Bilous, M., Fitzgibbons, P. and Hanna, W., 2013. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *Journal of clinical oncology*, *31*(31), pp.3997-4013.

Zulkepli, J., Eldabi, T. and Mustafee, N., 2012, December. Hybrid simulation for modelling large systems: an example of integrated care model. In *Simulation Conference (WSC), Proceedings of the 2012 Winter* (pp. 1-12). IEEE.