Faculty of Applied Sciences
Rajarata University of Sri Lanka

Research Pre-proposal

of

Individual Research Project – ICT 4609

for the degree of

B.Sc. (Honors) in Information Technology

(*Leave this page blank*)

Details of the Research Project

Research Title : An Efficient Breast Cancer Detection Algorithm Based on Artificial Neural Networks

(tentative)

Details of the student

Name : Kavindya Gimhani Rajapaksha

Registration ID : ICT/2016/2017/083

Index No. : 3946

Degree Program : BSc in Information Technology

Subject Combination : -

Department / Unit : Department of Computing

Details of Supervisor(s)

Name : Wiraj Udara Wickramaarachchi

Department / Unit : Department of Computing

Institute : Rajarata University of Sri Lanka

Contact Details : 0776329957 / wirajudara@as.rjt.ac.lk

Guideline to fill the research pre-proposal document

1. Research pre-proposal should be prepared using a word processor, clear and coherent.
2. The content of each section in research pre-proposal should be in detail and make the focal points stand out.
3. A student can have more than one supervisor. In such case, the details of all the supervisors should be included on the proposal document (the student may copy the same structure given in the details page).
4. The student can attach additional pages for section 1 and 2.
5. The student must follow the IEEE format to make the reference list.
6. This document should be printed on both sides of A4 papers (do not print with colors).
7. The students of the physical sciences department who are taking the course should change the course name and the degree name in the cover page.

Submission Guidelines for research proposal document

1. The student must submit this document along with the application for the 4^{th} year degree on or before the deadline.
2. The student must submit a soft copy of finalized research pre-proposal document to the supervisor, as a word processor document (.doc or .docx) on or before the deadline.
3. The supervisor should submit the finalized soft document to the department after the required corrections and with his/her approval which is symbolized by his/her signature in the section 3 of this document.
4. Meanwhile, the application and the hard copy of the pre-proposal document which is recommended and approved by the supervisor(s) (with signature(s)), must be submitted to the department before the deadline.

1. Introduction

1.1. Tentative title of the research project

An Efficient Breast Cancer Detection Algorithm Based on Artificial Neural Networks

1.2. Purpose and significance of the research study

Breast cancer is a cancer that develops from the breast tissues and arise in the lining cells of the ducts, lobules or in the glandular tissue of the breast. Over time these malignant tissues can progress and invade the surrounding breast tissues. They can further spread into the lymph nodes and then into the other organs of the body. Breast cancer occurs due to metastasis which is a growth of a secondary malignant at a distance from the primary site of the cancer can cause fatalities.

In 2020, there were 2.3 million women detected with breast cancer and 685 000 deaths was reported globally. Between 2015 to 2020, there were 7.8 million women alive who were diagnosed with breast cancer, making it the world's most dominant type of cancer [1].

Mortality rate due to breast cancer decreased slightly between 1930s through the 1970s due to early cancer detection. Mammography is one of the most common techniques that is used to detect cancer. During the recent past several research studies took place to automate the early diagnosis of breast cancer. The advancement of information technology provides image processing techniques for early identification of cancer tumors.

The objective of this research is to explore the currently available breast cancer detection algorithms which use image processing techniques and contribute to the enhancement of the accuracy of these algorithms by (i) increasing the sample size by using the same set of data, (ii) changing the training method and ensemble learning, (iii) changing the model parameters via hyper parameter tuning and (iv)using image processing techniques.

1.3. Research problem(s) / question(s)

1. How the existing breast cancer detection systems work and what is the current status in terms of efficiency?
2. How the feature extraction and preprocessing techniques affect the efficiency of breast cancer detection algorithms?
3. How does existing classification techniques for breast cancer detection work?

1.4. Expected research results and/or innovations

a) A comparative study of different classification techniques will provide insights about an efficient breast cancer detection algorithm.
b) A comparative study of feature extraction and preprocessing techniques will provide insights about an efficient breast cancer detection algorithm.
c) A comparative study of the efficiency of the existing breast cancer detection algorithms.
d) A comparative study of different data sets.

2. Summary of review of literatures: ***(The student must refer and cite 10 most related references. Show the list of references at the end of this section)***

There were many research studies being conducted to detect breast cancer using image processing and machine learning algorithms. However, there were very few research being done to use image processing techniques and machine learning algorithms combined. In a research study which was conducted by Mahendra G. Kanojia and Siby Abraham; Radial Basis Function neural network was used to detect breast cancer. In the research methodology of this study, the first step includes image processing methodologies for image enhancement and object localization. As the second step, feature extraction is conducted. In the third step, the Radial Basis Function Neural Network is used to classify the image as "benign" that is noncancerous or "malignant". In this study k-means clustering algorithm was used for the classification. Here, a statistical learning algorithm known as "*decision-with-rejection*" methodology was used. This methodology is semi-automatic, where the system generates a prediction for a confidence that is greater than the threshold and the lower confidence decisions are assigned to a human expert [2].

In another study, the two scientists Yadollahpour Ali and Shoghi Hamed explored different types of image processing techniques that are currently used for breast cancer detection. According to this study, misinterpretation of a lesion by the radiologist can lead to ~~a~~ larger number of false positive results. As a solution for this issue, computer aided detection (CAD) techniques were used to process and analyze mammogram images [9]. This helped the radiologists with identification and classification of masses. The study further discussed the common techniques that are used to detect masses in the mammograms. These techniques include, image preprocessing, image segmentation, feature extraction, feature selection, classification, and evaluation [3]. In their study, Ali and Hamed considers image preprocessing as one of the most significant practices in machine learning driven breast cancer detection. Image preprocessing is generally divided in to three categories namely denoising, enhancement of structure and enhancement of contrast. The most popular enhancement techniques that are used for mammograms today include, un-sharp masking (an image sharpening technique [4]), region-based enhancements [5], and optimal adoptive enhancement [6]. This research article also addresses the significant image segmentation methods such as thresholding, edge-based segmentation, region-based segmentation, clustering, classifier-based segmentation, and deformable model-based segmentation. Edge detection is another common technique that is used to detect masses in mammograms. This refers to algorithms which try to identify edges. The gray histogram and gradient based method are the two major edge-based algorithms that are used today [7]. Edge detectors consists of different operators such as Laplacian, and Sobel for detecting edge features. Image segmentation refers to the technique of partitioning a single image into multiple segments to make the image more meaningful and simpler to analyze. Global thresholding is the most common type of segmentation technique that is used to detect masses in images [8]. A mass or a tumor usually have a greater intensity compared to the surrounding tissues. This characteristic could be used to find the global threshold value. In a histogram, the regions with the abnormalities have extra peaks while the healthy region has only a single peak [9].

Clustering is the process of organizing objects into different groups based on their attributes. Similarly, images could also be grouped based on the similarity of the features, shapes, and textures. The most common clustering algorithms that are used presently are hard clustering, k-means clustering, fuzzy clustering etc. [9]
Feature extraction is another technique that could be used to detect masses in a mammography image. The extracted features are recognized according to the method of extraction and the image characteristics. Examples of the types of features extracted are,

- Features extracted from spatial grey level dependence matrix (co-occurrence matrix)
- Features extracted from the grey level run length (GLRL) matrix
- Features extracted from the grey-level difference (GLD) matrix
- Energy, entropy, and norm extracted from the wavelet transform sub-images [9]

The most common type of classifier used in recognizing masses in mammograms is the K-nearest neighbor (KNN) classifier which is used to recognize unknown patterns based on the similarity to known samples [9]. The artificial neural networks (ANNs) are non-parametric pattern recognition systems that can extract general rules by learning from real data or examples is another classifier that is used in breast cancer detection presently [9].

2.1 Analysis of existing works of the study area, including their development and existing problems

Reading mammography is a very challenging job for the radiologists as a misinterpretation of a lesion can lead to a larger number of false positive results. As a solution for this issue computer aided detection (CAD) techniques were used to process and analyze mammogram images. This helped the radiologists to identify masses and for classification. By using CAD, the accuracy of breast cancer detection was elevated to 90% from 80% [10].

The techniques in the field of computer aided mammography include image preprocessing, image segmentation techniques, feature extraction, feature selection, classification techniques etc. Further investigations into each of these steps are required to improve the overall performance of computer aided detection and diagnosis algorithms [11].

In a research study which was conducted using Radial Basis Function neural network to detect breast cancer by Mahendra G. Kanojia and Siby Abraham, the accuracy was calculated as 73% as the results of this study was produced based on an online breast cancer features database. The dataset which was obtained for this study had a minimum opacity and covariance [2].

In a similar research study, the time consumed by each image processing technique was evaluated. The techniques that were evaluated here are thresholding, edge-based and watershed methods. According to the results obtained from this research study it could be observed that although the threshold technique was the fastest, the other two techniques generated clear output images compared to the threshold method [12].

A comparative study of the different diagnosis techniques on mammographic images was conducted by a group of researchers. Here the characteristics and the limitations of each of the techniques were discussed. According to the study, digital mammography is very useful to identify calcification, and other ill-defined masses. A Further, the image segmentation and 3D structure analysis is useful in calculating the exact size and thickness of the tumors. It was also discovered that the accuracy of the enhanced watershed method to be almost 96% and that of MRF and HPACO algorithms to be 94.8% [13].

It is evident from the literature discussed above, that there were many research studies being done in the field of image processing to find the tumor. However, the accuracy rates of these studies varies between 75% to 92% and there are methods and techniques that can be explored to further to improve these accuracies rates [11]. Hence, my objective is to evaluate these existing techniques and enhance the effectiveness of these algorithms by means of changing model parameters, sample size or by changing the training method or by ensemble learning.

2.2 References *(please follow the IEEE referencing style)*

[1] WHO, "Breast cancer," *www.who.int*, Mar. 26, 2021. https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

[2] Kanojia MG and Abraham S, "Breast Cancer Detection Using RBF Neural Network," Accessed: Dec. 05, 2021. [Online]

[3] Mohanty AK, Champati PK, Swain SK, Lenka SK. A review on computer aided mammography for breast cancer diagnosis and classification using image mining methodology. International *Journal of Computer Science and Communication*. 2011; **2**(2):531-8.

[4] Chan H-P, Vyborny CJ, MacMAHON H, Metz CE, Doi K, Sickles EA. Digital Mammography: ROC Studies of the Effects of Pixel Size and Unsharp-Mask Filtering on the Detection of Subtle Microcalcifications. *Investigative Radiology*. 1987; **22**(7): 581-9.

[5] Morrow WM, Paranjape RB, Rangayyan RM, Desautels JEL. Region-based contrast enhancement of mammograms. *Medical Imaging, IEEE Transactions on*. 1992; **11**(3): 392-406.

[6] Dhawan AP, Buelloni G, Gordon R. Enhancement of mammographic features by optimal adaptive neighborhood image processing. *Medical Imaging, IEEE Transactions* on. 1986; **5**(1): 8-15.

[7] Kang W-X, Yang Q-Q, Liang R-P, editors. The comparative research on image segmentation algorithms. First International Workshop on Education Technology and Computer Science; 2009

[8] Brzakovic D, Luo X, Brzakovic P. An approach to automated detection of tumors in mammograms. *Medical Imaging, IEEE Transactions* on. 1990; **9**(3): 233-41

[9] Ali*Y and Hamed S, "Early Breast Cancer Detection using Mammogram Images: A Review of Image Processing Techniques," Mar. 10, 2015.

[10] Cheng H, Shi X, Min R, Hu L, Cai X, Du H. Approaches for automated detection and classification of masses in mammograms. *Pattern recognition*. 2006; **39**(4):646-68.

[11] Giri P and Saravanakumar K, "Breast Cancer Detection using Image Processing Techniques," *ORIENTAL JOURNAL OF COMPUTER SCIENCE & TECHNOLOGY*, vol. 10, no. 2, pp. 391–399, Apr. 2017, Accessed: Dec. 05, 2021. [Online].

[12] Alhadidi B, Zu`bi MH, and Suleiman HN, "Mammogram Breast Cancer Image Detection Using Image Processing Functions," *Information Technology Journal*, vol. 6, no. 2, pp. 217–221, Feb. 2007, doi: 10.3923/itj.2007.217.221.

[13] Wickramaarachchi WU, Kariapper RKAR, Haleem SLA, and Rukshana SLR, "Study on Mammographic Usage of Image Processing Techniques," Accessed: Dec. 06, 2021. [Online].

3. **Recommendation of supervisor(s) on the research problem and research pre-proposal** ***(This section should be filled by the supervisor(s). Supervisor(s) may consider the adequacy and scope of the research problem, quality and adequacy of the reviewed literature).***

   Comments (if any):

   I certify that, the student engaged continuously with me in developing the pre-proposal and, I am confident that he is adequately competent to continue this research.

   Signature(s) of Supervisor(s):

   Date: 03/01/2022

4. Research pre-proposal assessment team ***(this section should be filled by the department)***

Date of research pre-proposal presentation:

Panel members	**Name**	**Department / Institute**
Chair		
Member		
Member		
Member		
Member		

5. **Comments of the assessment team on the research** ***(This should be filled by the chair of the assessment panel. In case of revision or fail, needed revision in the pre-proposal or reasons to reject the pre-proposal should be mentioned here)***

The proposed research is adequate / not-adequate to full fill the requirement of the degree. And the context of the research is enough to award such a degree.

Signature of the panel chair	
Date	