Universiti Teknologi MARA

Feature Selection for Breast Cancer Diagnosis via Visualization

Izzah Khairina Bt Muhadi

Thesis submitted in fulfilment of the requirements for Bachelor of Computer Science (Hons.) Computer Science
Faculty of Computer and Mathematical Sciences

January 2019

ACKNOWLEDGEMENT

Alhamdulillah, praises and thanks to Allah because of His Almighty and His utmost blessings, I was able to finish this research within the time duration given. Firstly, my special thanks go to my supervisor, Pn Nor Ashikin Mohamad Kamal. for always correcting my mistakes and guiding me to make this report perfect. Without her help and her knowledge, it is impossible to complete this project. Thank you to my CSP600 lecturer, Dr Zainura bt Idrus and my CSP650 lecturer, Dr Marina Ismail for all the useful guidance in completing this research.

Special appreciation also goes to my beloved parents as without them I would not have been where I am today. There are not enough words to express their sacrifice, help and support towards me.

Last but not least, I would like to give my gratitude to my dearest friend.

ABSTRACT

Early diagnosis of breast cancer is important as it is one of the reason that causes death among women and men. Most diagnostic systems suffer the feature multiplicity problem. Some of these features are redundant and irrelevant to be used for breast cancer classification. Feature selection techniques have been used to find the most important features that are suitable to classify the type of breast cancer either malignant or benign. From the previous research work, it shows that visualization can contribute to feature selection. This project explores the feature selection through visualization as opposed to chi square filter feature selection technique. The visualization technique used for this project is a radial chart using d3.js library. Each feature of the data is the axis in radial chart and it will plot based on the value of the features. The radial chart used two different colours to differentiate between malignant and benign and the features are selected based on the features that are least overlap when the data being plotted. The new features that have been chosen from the visualization technique are used to classify the breast cancer type using K-Nearest Neighbour (KNN) classifier. To evaluate effectiveness of the proposed feature selection technique, the results are compared with the chi square filter feature selection technique using accuracy, specificity and sensitivity measurement. The results show that feature selection via visualization produce higher accuracy, specificity and sensitivity after being compared with chi square technique.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This chapter provides the background and rationale for the study. It also gives details of the significance of the differentiation between malignant and tumor over the Internet, the issues and problems that led to this research.

1.1 Background of Study

The uncontrolled growth of abnormal cells that can be anywhere in a body is called cancer (Hejmadi, 2010). The most commonly diagnosed cancer among woman is **breast cancer.** Breast cancer is the malignant tumours that forms in the cells of breast and originate in the breast tissue that grow erratically (Akay, 2009). According to American Cancer Society, the cells that turned a tumour can be displayed on an x-ray or a lump can be felt (DeSantis, Siege, & Jemal, 2018). If the cells grew into the tissues surrounding or spread to nearby areas of the body, the tumour is malignant that is a cancer.

Any part of breast can be the start of the breast cancer. There are case where it starts in the glands that produce breast milk and there are some start in the ducts where the milk is carried to the nipple. The symptoms and signs that someone have breast cancer are there is a lump in the breast or armpit, inverted nipple or they felt pain at their breast or have sore nipple (DeSantis et al., 2018). Breast cancer can be diagnosed through mammography, ultrasound or biopsy.

According to Yip, Taib, & Mohamed (2006), out of 20 women in Malaysia, one of them are suffering from breast cancer in their lifetime. Malaysia and others country around the world have higher percentage of patients that were diagnosed with breast cancer. In India, there are increasing number of patients that are diagnosed with breast cancer even in their thirties and forties (Kumar