VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI - 590018

Project Report

on

" FEATURE EXTRACTION AND CLASSIFICATION OF MICROCALCIFICATION CLUSTERS"

Submitted in partial fulfilment of the requirements for the VIII Semester Bachelor of Engineering

in

COMPUTER SCIENCE AND ENGINEERING

For the Academic Year

2019-2020

BY

PARTH SACHDEV	**1PE16CS108**
SANDHYA K S	**1PE16CS140**
UDDIP YALAMANCHILI	**1PE16CS170**
VRAJ RESHAMDALAL	**1PE16CS179**

Under the Guidance of

Prof. Sangeetha R

Assistant Professor, Dept. of CSE, PESIT-BSC

Department of Computer Science and Engineering

PESIT - BANGALORE SOUTH CAMPUS

Hosur Road, Bengaluru - 560100

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI - 590018

Project Report

on

" FEATURE EXTRACTION AND CLASSIFICATION OF MICROCALCIFICATION CLUSTERS"

Submitted in partial fulfilment of the requirements for the VIII Semester

Bachelor of Engineering

in

COMPUTER SCIENCE AND ENGINEERING

For the Academic Year

2019-2020

BY

PARTH SACHDEV	1PE16CS108
SANDHYA K S	1PE16CS140
UDDIP YALAMANCHILI	1PE16CS170
VRAJ RESHAMDALAL	1PE16CS179

Under the Guidance of

Prof. Sangeetha R

Assistant Professor, Dept. of CSE, PESIT-BSC

Department of Computer Science and Engineering

PESIT - BANGALORE SOUTH CAMPUS

Hosur Road, Bengaluru - 560100

PESIT - BANGALORE SOUTH CAMPUS

HOSUR ROAD, BENGALURU - 560100

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project work entitled ***"Feature Extraction And Classification Of Microcalcification Clusters"*** carried out by ***"Parth Sachdev, Sandhya K S, Uddip Yalamanchili, Vraj Reshamdalal*** bearing USN's ***1PE16CS108, 1PE16CS140, 1PE16CS170, 1PE16CS179"*** respectively in partial fulfillment for the award of Degree of Bachelors (Bachelors of Engineering) in Computer Science and Engineering of Visvesvaraya Technological University, Belagavi during the year 2019-2020. It is certified that all corrections/ suggestions indicated for internal assessment have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said Degree.

Signatures:

Signature of the Guide	Signature of the HOD	Signature of the Principal
Prof. Sangeetha R	**Dr. Sandesh B. J.**	**Dr. J Surya Prasad**
Assistant Prof,CSE	**HOD, CSE**	**Principal,PESIT-BSC**

External Viva

Name of the Examiners **Signature with Date**

1. ———————-

2. ———————-

ACKNOLEDGEMENT

The satisfaction and euhoria that accompany he successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned my effort with success.

We are indebted to our Guide, **Prof.Sangeetha R**, Assistant Professor, Department of Computer Science and Engineering, PESIT - Bangalore South Campus, who has not only coordinated our work but also given suggestions from time to time.

We are also extremely greatful to our Project Co-ordinatosr, **Dr. Ranjitha Kumari Dash**, **Prof. Shubha Raj K.B**, Assistant Professor, **Prof. Divya Prabha K.N**, Assistant Professor, Department of Computer Science and Engineering, PESIT - Bangalore South Campus, for their constant support and advice throughout the course of the preparation of this document.

We are greatly thankful to **Dr. Sandesh B J**, Professor and HOD,Department of Computer Science and Engineering, PESIT - Bangalore South Campus, for his able guidance , regular source of encouragement and assistance throughout this project.

We would like to express our immense gratitude to **Dr. J. Suryaprasad**, Director and Principal, PESIT - Bangalore South Campus, for providing us with excellent infrastructure to complete our project work.

We are greatfully acknowledge the help lent out to us by all faculty members of the Department of Computer Science and Engineering, PESIT - Bangalore South Campus, at all difficult times. We would also take this opportunity to thank our college management for the facilities provided during the course of the project.

Parth Sachdev
Sandhya K S
Uddip Yalamanchili
Vraj Reshamdalal

ABSTRACT

Breast cancer is the most common invasive cancer in women and the second main cause of cancer death in women after lung cancer. With breast cancer, early detection is key. The earlier the disease is diagnosed , the better the outcome with treatment.

This project aims in detecting the breast cancer at a very early microcalcification stage itself rather than at tumor stage using different efficient techniques of extraction of the features such as brightness, contrast, size, shape from microcalcification clusters and then classifying the microcalcification clusters into either benign or malignant cancer cells with the help of the extracted features.

Keywords: microcalcification, benign, malignant

Contents

List of Figures

Chapter 1

Introduction

1.1 Purpose

Breast cancer is the second leading cause of female cancer mortality; 16% of all cancer cases in women are of breast cancer. The breast cancer cases in developed countries are very high and it is a serious concern. Whereas in less developed countries the numbers of cases are continuously rising. The mortality rate is very high in developing countries as compared to developed countries. The low survival rates in developing countries may be due to lack of awareness and delayed detection. The detection of breast abnormalities at the initial stage is important to increase the survival rate since it restricts a woman from entering into further stages of breast cancer.

1.2 Scope

- The main objective is to identify presence of breast cancer at the microcalcification stage.
- Many features such as contrast,size,brigthness,shape and texture are drwan out from the microcalcification clusters employing different efficient feature extraction techniques.
- Classifying clusters as malignant or benign cancer cells with the help of extracted features.

1.3 Statement of the problem

A novel approach for classification of benign and malignant Microcalcification cluster.

1.4 Existing System

Research on breast cancer using digital image processing is not new but lack of proper methods for early detection at microcalcification stage is still a challenge to medical domain.Most of the research work done till now detects the breast cancer at tumor stage and are not accurate at 100% and leads to false positive or false negative results which are highly dangerous.

1.5 Proposed System

The proposed model uses a combination of different highly efficient techniques of digital imaging to extract different features from microcalcification clusters and then classifying clusters into either benign or malignant cells. The proposed system will be of high accuracy leading to true positive and true negative results.

1.6 Summary

This chapter gave a brief introduction on what exactly the proposed system is. It also covered the future scope and the demand of this product.It shows the main features overcoming the drawbacks of the existing system.

Chapter 2

Literature Survey

- Comaprision of Classifiers:(Competitive Study of Classification Techniques on Breast Cancer FNA Biopsy Data - Daniele Soria ,Jonathan M. Garibaldi,Elia Biganzoli,Ian O. Elli)

1. Decision tree algorithms: Frequently used decision tree algorithms are ID3,C4.5,and CART.CART (Classification and Regression Trees)provides better accuracy in classifying the breast cancer data sets than ID3,C4.5 algorithms. C4.5 Classifier : C4.5 is an algorithm used to generate a decision tree and these trees can be used for classification. C4.5 uses the concept of information gain to make a tree of classifactory decisions with respect to a previously chosen target classification. The output of the system is available as a symbolic rule base. The cases are for patterns that allow the classes to be discriminated. These patterns are then expressed as models, in the form of decision trees, which can be used to classify new cases.Since for the real world databases the decision trees become huge and diffcult to understand and interpret scrutinized.

2. Multilayer Perceptron Classifier: A multilayer perceptron is a forward artificial neural network model that maps sets of input data onto a set off appropriate output. It has distinctive characteristics:
   1. Model of neuron in the network includes a nonlinear activation function.
   2. The network consists one or more layers of hidden neurons that are not part of input and output of the network. Error back-propagation algorithm is used in this type of classifier

3. Nave Bayes Classifier: A nave bayes classifier is a simple probabilistic classifier based on applying Bayes theorem with strong independence assumptions It is based on two simplifying common assumptions: firstly, it assumes that the predictive attributes are conditionally independent given the class and sec-

ondly, the values of numeric attributes are normally distributed within each class. Bayes theorem is given as : p(C=c/X=x)= p(C=c)p(X=x/C=c)/p(X=x).

4. Support Vector Machine : SVM is a technique based on the statistical learning theory.It seperates two classes by determining the linear classifier that maximizes the margin . According to the paper on comparitive study of classification by Haowen You and Gearge Rumbe, accuracy of this classifier remained constant at 62.74% and progressively better predictions above 90 percent were determined. Initially in this method SVM was trained by varying gamma values and range was selected and classifier was evaluated including functions like polynomial, sigmoid and radial bias function

- Automatic feature extraction for breast density segmentation andclassification

The breast cancer detection mainly involves three major steps.
Step 1 : Involves the identification of ROI from the mammograms by the method of image segmentation. This enables to find the region with calcium deposits as the region with a tumor.
Step 2 : Involves the dataset creation containing the features required for the study of cancer by the process of feature extraction.
Feature selection helps in identifying the attributes that contribute to our study. There are different types of features extracted from mammograms that include statistical, structural, and textural features.
Step 3 : classification to segregate mammograms as benign, malignant, and normal tissues. The KNN and SVM classifiers were implemented to classify images based on the features extracted. It was proved that SVM provided the maximum accuracy of classification comparatively with 88.67% accuracy. This proves that the SVM is a better classifier for classifying extracted features.

- On Breast Cancer Detection: An Application of MachineLearning Algorithms on the Wisconsin Diagnostic Dataset

Preprocessing : $z = \frac{(X-\mu)}{\sigma}$
All presented ML algorithms exhibited high performance on the binary classification of breast cancer, i.e. determining whether benign tumor or malignant tumor. Consequently, the statistical measures on the classification problem were also satisfactory.

Chapter 3

Hardware and Software Requirements Specification

3.1 Software Requirements Specifications

Requirement specification is the movement of interpreting the data assembled amid investigation into prerequisite report.
Software requirements specifications are the detailed enlisting of all necessary requirements that arisein the project. The aim of having these requirements is to gain an idea of how the project is to be implemented and what is to be expected as a result of the project. The sections in this chapter deal with the various kinds of software, hardware and other functional and non functional requirements of the project. A brief description of the various users of the system is also mentioned.

3.1.1 Operating Environment

This section gives a brief about the hardware and software prerequisites for the project.
Hardware Requirements

- **Processor :**2GHz or faster processor.
- **RAM :** 2 GB(64bit).
- **Storage :** 250GB of available hard disk space.
- Other general hardwares such as a mouse and keyboard for inputs and a monitor for display.

Software Requirements

- **Operating System :** Ubuntu 14.04 and above or Windows 7 and above
- **Programming Languages :**Python
- **Documentations :**Overleaf

3.1.2 Functional Requirements

Functional requirements are a formal way of expressing the expected services of a project. We have identified the functional requirements for our project as follows:

- The system should be able to gather data from datasets.
- The system should be able to check for correctness of data extracted.
- The system should be able to efficiently extract features from the datasets.
- The system should be able to have the capacity to decide the contribution of each attribute towards the decision made by the predictor.
- The system should be able to classify the examples accurately.

3.1.3 Non-Functional Requirements

Non functional requirements are the various capabilities offered by the system. These have nothing to do with the expected results, but focus on how well the results are achieved.

- Reliability :The subsystem gives results with an high accuracy score. The systems are highly accurate subject to the proper working of each component involved.
- Scalability: The system should be able to run over huge datasets as well as small datasets and should be able to produce accurate results in both the cases
- Performance : The system will yield quick results for new queries.
- Portability : This is required when the computer, which is facilitating the framework stalls out because of few issues, which requires the system to be taken to another computer.
- Re-usability : The degree to which existing applications can be reused in new application. The predicted output could be reused in many fields.

3.1.4 User Characteristics

There is only one type of user associated with the system:

- Medical Personnel: He/She may use this system to detect the disease at the early stage which can be useful for providing proper treatment to the patients.

3.1.5 Applications

Medical domains: This system is used in medical domain for the diagnosis of breast cancer at the very early stage and will be useful in providing the proper treatment.

3.1.6 Summary

This chapter discussed the basic software and hardware requirements. More importantly it discusses the functional and non-functional requirements. Dept.

Chapter 4

System Design

High-Level Design

This section mainly covers the design technique of the entire system.:

- The Feature extraction phase and classifier

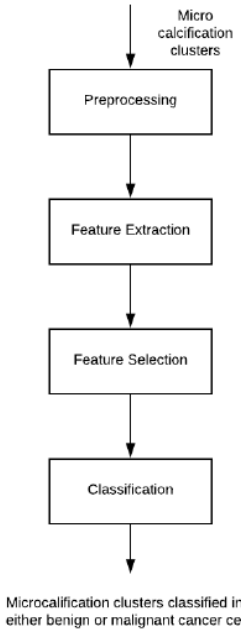

Figure 4.1: High-level design

4.1 Design Approach

Here are two methodologies for software designing:

- Top-down Design:It takes the entire programming framework as one entity and after that disintegrates it to accomplish in excess of one subsystem or some components based on few attributes.

- Bottom-up Design: The model begins with most particular and essential components. It accedes with making more elevated amount out of subsystems by utilizing essential or lower level

4.2 System Architecture

System architecture is the conceptual model that define the structure, behavior and views of a system. It is a formal description and representation of a system, the basic structure of proposed system is depicted in the below figure

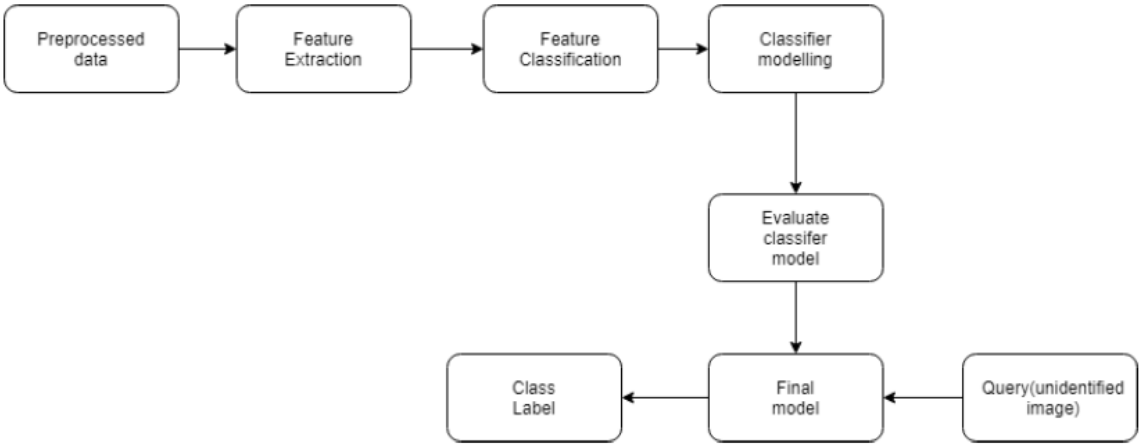

Figure 4.2: System Architecture

4.3 Data flow Diagram

Data Flow Diagram is the starting point of the design phase that functionally decomposes the requirement specification. A DFD consists of a series of bubbles joined by lines. The bubbles represent data transformation and the lines represent data flows in the system. A DFD describes what data flow rather than how they are processed, so it does not include hardware, software and data structure.
A Data Flow Diagram is a graphical representation of the "flow" of data through an information system. DFD's can also be used for the visualization of data processing (Structured Design). A data flow design is a significant modeling technique for analyzing and constructing information processes.A DFD can be referred to as a Process Model.

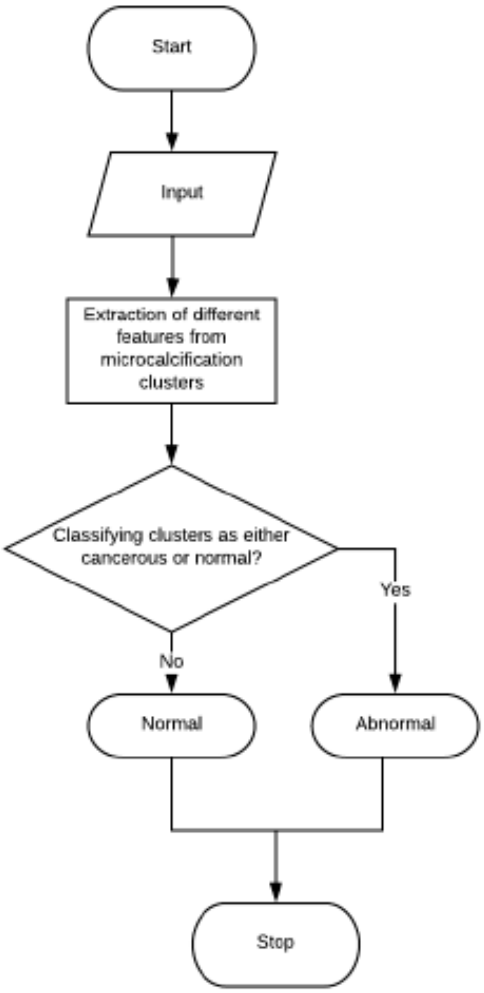

Figure 4.3: Data Flow Diagram

Chapter 5

Result Analysis

5.1 Expected Outcome

Classifies the test data into malignant or benign

The final feature set is constructed which is fed as input to the classification algorithm which classifies the the test data into malignant or benign using selected features.

References

[1] *Global detection approach for clustered microcalcifications in mammograms using a deep learning network.*; Juan Wang , Robert M Nishikawa, Yongyi Yang(2017)

[2] *Automatic Classification of Clustered Mirocalcification by a Multiple Expert Systems*; S. De Vito, F Tortorella , M Vento

[3] *On Breast cancer detection : An application of Machine learning algorithms on the Wisconsin diagnostic dataset.*; Abien Fred M.Agarap

[4] *Early detection of breast cancer using machine learning techniques*; Babak Bashari Rad

[5] *Analysis of Machine learning techniques applied to the classification of masses and microcalcification clusters in breast cancer computer-Aided detection.*; Jose L.Hernandez,Jesus S.Cepeda, Camila Castro

[6] *Classification of microcalcification in digital mammograms for the diagnosis of breast cancer*; Osamu Tsuji

[7] *Automatic feature Extraction for breast debsity segmentation and classification*; Ashwathy K Cherian, Poovammal E,Malathy C (2017)