

THESIS DECLARATION

I. Thesis Status Declaration Form

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signature:

Name: AHMAD BUKHARI BIN ABDULLAH

ID Number: CB11102

Date:

II. Thesis submitted in fulfillment of the requirements for the awards of degree

This thesis is being submitted in partial fulfillment of the requirements for the Bachelor Degree of Software Engineering at University Malaysia Pahang.

Signature:

Name: AHMAD BUKHARI BIN ABDULLAH

ID Number: CB11102

Date:

III. Student Declaration

I hereby declare that the work in this thesis is my own for quotations and summaries which have been duly acknowledge. The thesis has not been accepted for any degree and is not concurrently submitted for award of other degree.

Signature:

Name: AHMAD BUKHARI BIN ABDULLAH

ID Number: CB11102

Date:

IV. Supervisor Declaration

I hereby declare that I have checked this thesis and my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Software Engineering.

Signature:

Name of Supervisor: AZLINA BINTI ZAINUDDIN

Position:

Date:

ACKNOWLEDGEMENT

First and foremost, I have to thank my Undergraduate Project supervisors, Azlina binti Zainuddin. Throughout this semester, I went to Faculty of System Computer for several times to have a discussion with my supervisor. Despite from being busy she willing to sit together with me in helping me to finish this study. Without her assistance and dedicated involvement in every step throughout the process, this Study would have never been accomplished. I would like to thank you very much for your support and understanding over these past Years.

Not to forget to my Personal Advisor, Balsam binti Abdul Jabbar Mustafa for continuous support since the first time I come to University Malaysia Pahang. I am truly grateful for her progressive vision about my training in software engineering. Throughout these past year she had help me not only in my PSM but also in other subject related to Software Engineering. With her guidance I have get the view of what I will become after I graduated from this University.

Most importantly, none of this could have happened without my family. My mother, who offered her encouragement through phone calls for every week despite my own limited devotion to correspondence. With her own brand of humour, my Mother, Rokiah binti Hamid has been kind and supportive to me over not just for the last several years but since the beginning of my life. To my parents, my sister and my brother it would be an understatement to say that, as a family, we have experienced some ups and downs in the past three years. Every time I was ready to quit, you did not let me and I am forever grateful. This dissertation stands as a testament to your unconditional love and encouragement.

Abstract

These studies are being conducted to determine the most suitable Artificial Intelligent Technique to be implement in software Breast Cancer Detector (CBR). Breast Cancer Detector problem is to compare similarity of the new case to the old case which is have more than hundred record case. This is to ensure the less take time to compare one by one over the hundred case to a new case. The objective of this project is to develop the prototype to do the comparism of a new case with existing case of the breast cancer. Case Base Reasoning (CBR) is capable of solving the measurement of similarity and less take time to find the highest similarity. CBR consist four phase to be done to solve the similarity measurement. The first phase is retrieve that is problem (new case) is retrieved. The second phase is reuse that is reuse the solved case and calculate to find the suggested solution or it called the highest similarity in percentage. The third phase is revise which is process to confirm solution (teste or repaired case). The last phase is retain. The machine learn in other mean the machine save the new case if the new case does not have the highest similarity and the doctor should do the phisycal check. The calculation that used to calculate the similarity measure is called Feature-Based Similarity Measure algorithm.

Abstrak

Kajian-kajian ini dijalankan untuk menentukan Teknik Pintar (AI) tiruan yang paling sesuai untuk melaksanakan dalam perisian Breast Cancer Detector (CBR). Masalah Breast Cancer Detector (CBR) adalah untuk membandingkan persamaan kes baru dengan kes lama yang mempunyai lebih dari seratus kes rekod. Ini adalah untuk memastikan persamaan untuk membandingkan satu demi satu ke atas kes itu kepada ratusan kes baru. Objektif projek ini adalah untuk membangunkan prototaip untuk melakukan perkara yang mengira persamaan sesuatu kes baru dengan kes yang sedia ada (Kanser Payudara). Case-Based Reasoning (CBR) mampu menyelesaikan pengukuran persamaan dan kurang mengambil masa untuk mencari persamaan yang paling tinggi. CBR terdiri empat fasa yang perlu dilakukan untuk menyelesaikan pengukuran persamaan. Fasa pertama adalah mendapatkan semula iaitu masalah (perkara baru) adalah "retrieved". Fasa kedua ialah "reuse" yang menggunakan semula kes itu diselesaikan dan mengira untuk mencari penyelesaian yang dicadangkan atau ia dipanggil persamaan yang paling tinggi dalam peratusan. Fasa ketiga "revise" yang merupakan proses untuk mengesahkan penyelesaian (uji atau kes dibaiki). Fasa terakhir ialah "retain". Mesin belajar atau dipanggil mesin menyimpan kes baru jika mana-mana kes yang baru tidak mempunyai persamaan yang paling tinggi dan doktor perlu buat cek secara fizikal. Pengiraan yang digunakan untuk mengira persamaan dipanggil "Feature-Based Similarity Measurement algorithm".

Table of Contents

Thesis Declaration.....	i
i. Thesis Status Declaration Form	i
ii. Thesis.....	ii
iii. Student Declaration	iii
iv. Supervisor Declaration	iv
Acknowledgement.....	v
Abstract.....	vi
Abstrak	vii
Chapter 1(Introduction).....	3
1.1 Background	4
1.2 Problem Statement	4
1.3 Aim and objectives of the project.....	4
1.4 Scope	5
Chapter 2(Literature Review)	6
2.1 Literature Review	6
2.1.1 Introduction of Case-Based Reasoning.....	6
2.2 Case-Based Reasoning (CBR)	7
2.3 Advantages And Disadvantages Of Case-Based Reasoning Technique	8
2.3.1 Advantages	9
2.3.2 Disadvantages.....	9
2.4 Existing Systems.....	10
2.5 The Process Of Case-Based Reasoning.....	12
2.6 Breast Cancer	17
2.6.1 Data Set Breast Cancer Wisconsin (Original)	17
2.6.2 Attribute Information.....	18
2.6.2.1 Benign Tumors.....	19
2.6.2.2 Malignant Tumors	19
2.6.2.3 Definition Attribute of Cancer.....	19
Chapter 3(Methodology).....	21

3.1 Introduction.....	21
3.2 Methodology	23
3.2.1.1 Proposed Methodology	23
3.2.1.2 Steps Involves In Methodology(Analysis)	23
3.2.1.3 Proposed Technique	25
3.2.2 Hardware & Software	26
3.3 Gantt Chart	26
Chapter 4(Design and Implementation)	27
4.1 Introduction.....	27
4.2 Result Analysis	27
4.3 Constraints	28
4.3.1 Development constraints	28
4.3.2 System constraints	28
4.4 Developing the Software.....	29
4.5 Result of the Application	31
4.6 Coding and Explanation.....	35
4.6.1 Coding CBR – Featured Based Reasoning	35
4.6.2 Read Data Set from Database.....	38
4.7 System Limitation	41
4.8 Advantage of this Application	41
4.9 Future Enhancement	42
Chapter 5(Conclusions).....	43
5.1 Introduction.....	43
5.2 Summary of Literature Review	43
5.3 Summary of Methodoly	43
References.....	44
APPENDIX.....	46
APPENDIX A : PROJECT GANTT CHART	47
APPENDIX B : PSM POSTER	48

Chapter 1

Introduction

1.1 Background

I want to build the project is to ease specialist software to detect breast cancer patients more quickly, reduce calculation error and comforting doctor for not looking at all the old records of patients who have suffered from cancer for the conclusion that the patient suffers from cancer at what level or stage.

In the information, an early sign of breast cancer occurs when swelling or lump in the breast, bloody fluid out of the nipple, enlargement of glands in the armpits, wrinkles in the skin of the breasts. It starts in the cell of the breast by the group of cancer cells (malignant tumor)[1].

Breast cancer is most dangerous disease and it is the second leading cause of death among women. All human, both genders are born with some breast tissue and cells that have the possibility to get cancer breast. Male are rare to get breast cancer with only below 2500 diagnoses each year[2]. Even though most people exactly who produce bust most cancers will never be in a position to pinpoint just one certain lead to, scientists have discovered significantly concerning danger aspects that will show any stronger chances pertaining to most cancers.

There are many facts and myths that have been documented. Among them is smoking, family history, genetics, personal health history, early menstruation, late menopause after age 55, the use of agents, anti sweat result of breast cancer, all lumps are cancerous, the bigger breast size, the higher the risk, mammograms cause cancer, miscarriage induced causes of breast cancer and breast implants cause cancer[1][2].

Breast cancer tumors can be categorized by the size, type of cells, and the characteristics that fuel its growth. Breast cancer can be detected through the nine attributes which should be identified by a specialist, through inspection, that is the

clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses. After the inspection is done, the experts can confirm the patient's cancer is malignant or benign level. From the result of inspections carried out by a specialist, doctor selecting a scale from 1 to 10 for setting the level of any examination result. Collect all examination results of all nine attributes result compared with previous patients that have been recorded. Through these records, the calculation can be done through software that I will build. This software will use the artificial intelligence technique that is case base reasoning. Case-based reasoning (CBR) means using old experiences to understand and solve new problems. A case-based reasoning solves new problem by using or adapting solution that were used to solve old problems. A case-based reasoning is an AI technique that imitates how human make a decision. In CBR, new problems are solved by recalling from a previously solved problem which are stored in the case-base[4][5].

1.2 Problem Statements

There are five problem that can be solve by develop the system use the case-based reasoning method, which is first is the possibility of error calculations scale after examining the patient can be fixed, second is speed the time doctor to make a thorough examination, third is speed up time when the doctor compared the new patient records with an old patient records that who have experienced similar situations and last but not least is the wasting of existing data or record can be reduced.

1.3 Aim and objectives of the project

This project aims to develop an application that can provide information about interesting places in Kuantan and can navigate tourists to go to their selected places.

The objectives of the research are to:

- 1) To study case based reasoning artificial intelligent algorithm to apply in breast cancer detector software.
- 2) To develop a tool to process the similarity measurement and make the compares from the new record with previous record of the patient.
- 3) To evaluate the tool for improvement of detection of the cancer patients history.

1.4 Scope

- 1) This project concern on patients who related to the cancer detection focused on breast cancer.
- 2) This project also concern on doctor specialized in breast cancer.
- 3) The developed tool will use CBR calculation only.
- 4) Record of patient will use from Dr. William H. Wolberg (physician) from University of Wisconsin Hospitals, Madison, Wisconsin, USA dataset.
([https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Original\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original)))

Chapter 2

Literature Review

2.1 Literature Review.

Breast Cancer Detector Software -CBR(BCDS) created for use of doctors or expertise in the field of cancer breast. It used to detect the scale of the level of the cancer that the patient that has suffered from cancer. After the doctor or expert in cancer breast give the scale to attribute that related the cell of the cancer. Software BCDS will take the charge to give the final answer (beginning/danger/incurable). The scale of the cell that the doctor or expert should make their decision are the clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses. This system will use the artificial intelligence (AI) technique. The technique that will be used is case-based reasoning (CBR)[5][7].

2.1.1 Introduction of Case-Based Reasoning

Case-based reasoning (CBR), broadly construed, is the process of solving new problems based on the solutions of similar past problems. AI technique that tries to model/imitate human reasoning (decision making) and thinking. It has been argued that case-based reasoning is not only a powerful method for computer reasoning, but also a pervasive behavior in everyday human problem solving or more radically, that all reasoning is based on past cases personally experienced. Why it's called artificial intelligence technique? It is because the technique itself likes the human solving method. This technique used (adapting solution) the experienced problem that stored in the system to solve the problem that occurs or make the decision to solve. This technique also will learn itself (machine learning) when we used this method in the software. Scientific discipline concerned with the design and development of algorithms that allow computers to learn based on empirical data, such as from databases[6][7].

Case-based reasoning (CBR) has four processes that used to solve the problem (CBR cycle process). First is retrieving the similar problem from the database to solving it. Second, reuse the experience in the context of the new situation. That is complete or partial reuse, or adapt according to the differences. Third is revising process. This process is for tested and repaired case. It tests the new solution in the real world (or simulation) and if necessary. Fourth is retaining process that is after the solution has been successfully adapted to the target problem, store the resulting experience as a new case in the database (learned case or adapt case)[4][5].

Case-based reasoning (CBR) has used the similarity measure technique. Similarity measure is used in solving the problem and reasoning to match a previous experience or cases (case-base). This technique is the most important thing in the CBR software system. There are two types of similarity technique in similarity measure, local similarity and global similarity[6][7]. Local similarity used to compute the similarity between queries (new problem solving) and case attributes values.

2.2 Case-Based Reasoning (CBR)

Case-Based Reasoning (CBR) are one of the technique artificial intelligent. This technique is call as artificial intelligent is because of their action is like a human thinking. The comparable technique is just like the human based on the calculation. CBR works by compare the new data and existing data to find the result of the case. CBR use the similarity measured calculation to find the result of the new data[6].

There are four similarity measure calculation :

1. Distance-Based Similarity Measure
2. Probability Similarity Measure
3. Feature-Based Similarity Measure
4. Classification Similarity Measure

Feature-Based Similarity Measures:

Partly in response to empirical evidence against the distance axioms, Tversky (1977) proposed that perceived similarity is the result of a feature-matching process that differentially weights common and distinct stimulus features. Let $g(A \cap B)$ denote the salience of the features that are common to stimuli A and B and let $g(A - B)$ denote the salience of the features that are unique to stimulus A. Then Tversky's (1977) feature contrast model proposes that the similarity of stimulus A to stimulus B is equal to

$$s(A, B) = \alpha g(A \cap B) - \beta g(A - B) - \gamma g(B - A),$$

where α , β , and γ are constants that might vary across individuals, context, and instructions. According to this model, features in common increase similarity, whereas features that are unique to one stimulus decrease similarity. One advantage of the feature contrast model is that it can account for violations in any of the distance axioms.

2.3 Advantages And Disadvantages Of Case-Based Reasoning Technique

2.3.1 Advantages

i) Reduces the knowledge acquisition effort – Knowledge-Based System

Domain expert is unwilling to participate in the knowledge eliciting process due lack of understanding of the technology (technology gap).

The knowledge engineer finds difficulty in communicating with the domain expert/experts because the knowledge engineer, initially has less domain knowledge than the domain expert (knowledge gap).

Correctness of the rule-base developed can be questionable due to the knowledge gap between knowledge engineer and domain expert.

ii) White-box learning – explanation facilities

Explanation is a must to every intelligence system. It is a way how a user (human) can trust the reasoning done by the system. The need of the explanation facilities is

critical since the reasoning will be used in assisting the decision making process, giving advice or other contexts with potentially significant impact.

CBR stores old experiences or cases in a memory known as a case - base. This enables the algorithm to recall previous cases to solve a new problem. This process is called retrieval.

iii) Incremental learning

Incremental learning is an ability of a learning algorithm to incorporate new knowledge into existing knowledge to suite the available data. Learning new data without forgetting prior knowledge is an issue in learning algorithm.

iv) The closest match to human reasoning

Closer to human thinking and solving problem. Use case base reasoning is a technique or the way people think. Identifying the problem through experiences that have passed and if no experience, it will be a new experience for future use.

v) Makes use of existing data, (example: database)

Waste of old records of cancer patients do not occur. Old records can be used to compare the similarity of new record the patient's condition critical level for new patients.

2.3.2 Disadvantages

i) Lazy learning algorithm

Generalization the training data is delayed until a query is made to the system. Typically slower to evaluate the use of manpower, this case people are lazy to learn the techniques of calculation global similarity and local similarity. Large space requirement to store the entire training dataset (Not suitable for large data)

2.4 Existing Systems

i) **eXiT*CBR. v2: Distributed case-based reasoning tool for medical prognosis[10].**

The tool enables the health care collaboration practice to be mapped in cases where different doctors share their expertise, for example, or where medical committee composed of specialists from different fields work together to achieve a final prognosis. Each agent with a different piece of knowledge classifies the given cases through metrics designed for this purpose. Since multiple solutions for the same case are useless, agents collaborate among themselves in order to achieve a final decision through a coordinated schema. For this purpose, the tool provides a weighted voting schema and an evolutionary algorithm (genetic algorithm) to learn robust weights. Moreover, to test the experiments, the tool includes stratified cross-validation methods which take the collaborative environment into account. In this paper the different collaborative facilities offered by the tool are described. A sample usage of the tool is also provided.

Highlights

- Expand an existing medical focused tool to help physicians to design Case-based Reasoning experiments (CBR).
- Cooperative Multi Agent System (MAS) technology is used to improve CBR.
- The cooperation of different CBR agents improves the quality of the diagnosis.

ii) **Ontology-supported case-based reasoning approach for intelligent m-Government emergency response services. [11]**

There is a critical need to develop a mobile-based emergency response system (MERS) to help reduce risks in emergency situations. Existing systems only provide short message service (SMS) notifications, and the decision support is weak, especially in man-made disaster situations. This paper presents a MERS ontology-supported case-based reasoning (OS-CBR) method, with implementation, to support emergency decision makers to effectively respond to emergencies. The advantages of the OS-CBR approach is that it builds a case retrieving process, which provides a more convenient system for decision support based on knowledge from, and solutions provided for past

disaster events. The OS-CBR approach includes a set of algorithms that have been successfully implemented in four components: data acquisition; ontology; knowledge base; and reasoning; as a sub-system of the MERS framework. A set of experiments and case studies validated the OS-CBR approach and application, and demonstrate its efficiency.

Highlights

- Developed a mobile-based emergency response system to handle disaster situations
- Developed an extraction, intelligent aggregation algorithm to fuse emergency information
- Developed a case retrieval algorithm and a case adaptation algorithm using fuzzy sets
- Developed an OS-CBR approach to support emergency decision making
- Implemented the OS-CBR as a subsystem of the MERS framework

iii) Enabling the use of hereditary information from pedigree tools in medical knowledge-based systems. [10]

The use of family information is a key issue dealing with inheritance illnesses. This kind of information use to come in the form of pedigree files, which contain structured information as tree or graphs, which explains the family relationships. Knowledge-based systems should incorporate the information gathered by pedigree tools to assess medical decision making. In this paper, we propose a method to achieve such a goal, which consists in the definition of new indicators, and methods and rules to compute them from family trees. The method is illustrated with several case studies. We provide information about its implementation and integration on a case-based reasoning tool. The method has been experimentally tested with breast cancer diagnosis data. The results show the feasibility of our methodology.

Highlights

- Risk assessment of developing an inherited illness using pedigree information.

- Methods, structured and statistic indices are defined for knowledge-based systems.
- Real information from a breast cancer database is used for experimentation.
- Implemented and integrated into eXiT*CBR experimentation framework as a plugin.

2.5 The Process Of Case-Based Reasoning[12]

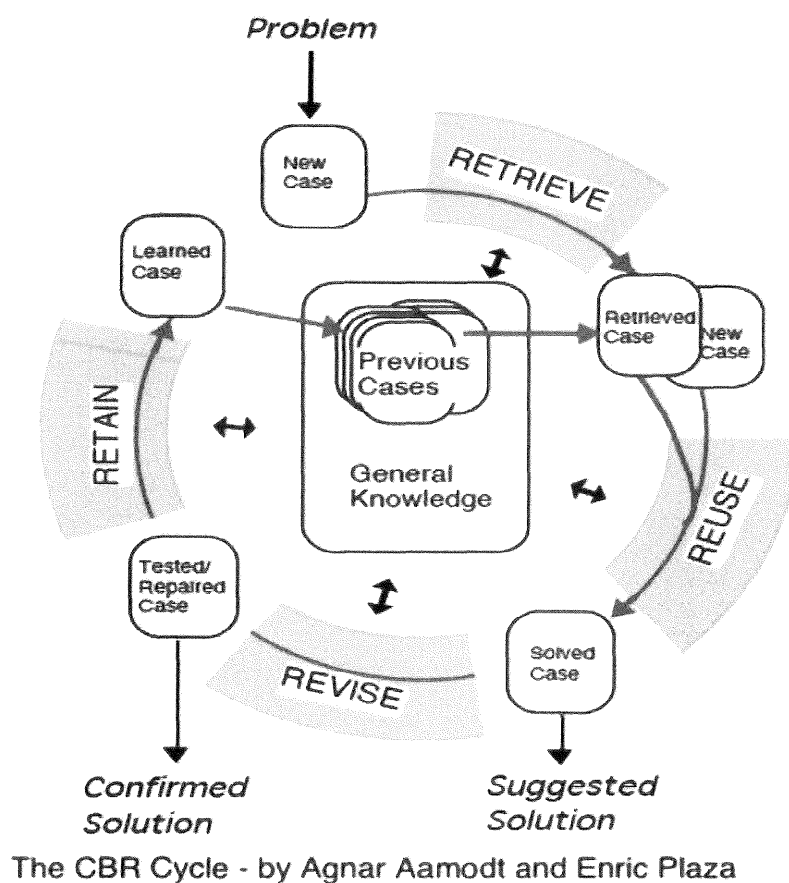


Figure 1: Case-Based Reasoning technique

I) RETRIVE

What is Case Retrieval?

- Usually regarded as the most important in CBR cycle
- Various approaches have been developed to retrieve similar cases in CBR

- Similarity measure is used in the CBR retrieval process
- Poor reasoning resulted from the poor similarity measure

What is Similarity measure?

- Similarity measure is used in problem solving and reasoning to match a previous experience/case (case-base) with the new unseen problem to find a solution.
- It is the heart of the CBR system!
- Can be called as case matching in CBR – match new case with the previous cases from the case base to find a solution.
- Purpose of similarity:
 - o Select cases that can be adapted easily to the current problem
 - o Select cases that have (nearly) the same solution than the current problem
- Basic assumption: similar problems have similar solutions
- Goal of similarity modeling: provide a good approximation

There are two types of similarity:

- Local Similarity
 - o Used to compute the similarity between the query (new problem) and case attributes values – feature level
 - o Two types of local similarity formula (Numerical/Continuous value and Discrete value)
- Global Similarity
 - o Global similarity is a build up from a number of local similarity functions. It is a weighted sum of the local similarity – case/object level
 - o After a set of local similarities have been calculated for each feature in the case, a global similarity will be calculated.
 - o Global similarity provides a case-matching behavior using the global similarity calculation to find the relationship between the two cases.

II) REUSE

Different option available:-

- No modification of the solution: just copy
- Manual/interactive solution adaptation by the user
- Automatic solution adaptation
 - o Transformational Analogy: transformation of the solution
 - o Derivational Analogy: replay of the problem solving trace
 - o Compositional adaptation: combine several cases to a single solution

III) REVISE

Verification & Correct Solution

- Revise phase: little attention in CBR research today
 - o No revise phase
 - o Verification of the solution by computer simulation
 - o Verification / evaluation of the solution in the real world
- Criteria for revision
 - o Correctness of the solution
 - o Quality of the solution
 - o Other, e.g., user preferences

IV) RETAIN

Learning from Solve Problem

- What can be learned:
 - o New experience (new case)
 - o Improved similarity assessment, importance of features

- Organization/indexing of the case base to improve efficiency
 - Knowledge for solution adaptation
 - Forgetting cases, e.g., for efficiency or because out-of-date
- Methods:
- Storing cases in the case-base
 - Deleting cases from the case-base

The Calculation case based reasoning:

Local Similarity

a = new feature

b = previous feature

range = value different between upper boundry and lower boundry of the set

Equation :

$$\mathit{similarity}(a, b) = 1 - \frac{|a - b|}{\mathit{range}}$$

Global Similarity**A = new case****B = previous case****a = new feature****b = previous feature****p = number of attribute****i = interation** **w_i = weight attribute i $\sum_{i=1}^p w_{i=1}$** **$similarity_i$ = local similarity calculate for attribute i****Equation :**

$$similarity(A, B) = \frac{1}{\sum w_i} \cdot \sum_{i=1}^p w_i \cdot similarity_i(a, b)$$

2.6 Breast Cancer

Breast cancer is a group of cancer cells (malignant tumor) that starts in the cells of the breast. Breast cancer is the second leading cause of death among women. All people, whether male or female, are born with some breast cells and tissue that have the possibility to develop into cancer. However, breast cancer in men is rare, with only about 2,190 diagnoses each year. Although most people who develop breast cancer will not be able to pinpoint one specific cause, scientists have learned much about risk factors that may indicate a stronger likelihood for cancer. Breast cancer tumors can be categorized by the size, type of cells, and the characteristics that fuel its growth[2][3].

2.6.1 Data Set Breast Cancer Wisconsin (Original)[1]

Data set that will be used in the prototype software defined below:

Data Set Characteristics:	Multivariate	Number of Instances:	699
Attribute Characteristics:	Integer	Number of Attributes:	10
Associated Tasks:	Classification	Date Donated:	15 / 07 / 1992

Table 1: Prototype Software Defined

Source:

Creator

Dr. William H. Wolberg (physician)

University of Wisconsin Hospitals

Madison, Wisconsin, USA

2.6.2 Attribute Information:

Attribute	Domain
1. Sample code number:	id number
2. Clump Thickness:	1 – 10
3. Uniformity of Cell Size:	1 – 10
4. Uniformity of Cell Shape:	1 – 10
5. Marginal Adhesion:	1 – 10
6. Single Epithelial Cell Size:	1 – 10
7. Bare Nuclei:	1 – 10
8. Bland Chromatin:	1 – 10
9. Normal Nucleoli:	1 - 10
10. Mitoses:	1 - 10
11. Class:	(2 for benign, 4 for malignant)

Table 2: Attribute and Levels

Class distribution:

Benign: 458 (65.5%)

Malignant: 241 (34.5%)

2.6.2.1 Benign Tumors[1]

When a tumor is diagnosed as benign, doctors will usually leave it alone rather than remove it. Even though these tumors are not generally aggressive toward surrounding tissue, occasionally they may continue to grow, pressing on organs and causing pain or other problems. In these situations, the tumor is removed, allowing pain or complications to subside.

2.6.2.2 Malignant Tumors[1]

Malignant tumors are cancerous and aggressive because they invade and damage surrounding tissue. When a tumor is suspected to be malignant, the doctor will perform a biopsy to determine the severity or aggressiveness of the tumor.

2.6.2.3 Definition Attribute of Cancer[2][3]

- Clump thickness: Benign cells tend to be grouped in monolayers, while cancerous cells are often grouped in multilayers.
- Uniformity of cell size/shape: Cancer cells tend to vary in size and shape. That is why these parameters are valuable in determining whether the cells are cancerous or not.
- Marginal adhesion: Normal cells tend to stick together. Cancer cells tends to loos this ability. So loss of adhesion is a sign of malignancy.
- Single epithelial cell size: Is related to the uniformity mentioned above. Epithelial cells that are significantly enlarged may be a malignant cell.
- Bare nuclei: This is a term used for nuclei that is not surrounded by cytoplasm (the rest of the cell). Those are typically seen in benign tumours.
- Bland Chromatin: Describes a uniform "texture" of the nucleus seen in benign cells. In cancer cells the chromatin tend to be more coarse.
- Normal nucleoli: Nucleoli are small structures seen in the nucleus. In normal cells the nucleolus is usually very small if visible at all. In cancer cells the nucleoli become more prominent, and sometimes there are more of them.

- Mitosis: Cell division that results in two daughter cells each having the same number and kind of chromosomes as the parent nucleus, typical of ordinary tissue growth. In simple word is the cell is divided into another cell but it similar like the first one before divide.

Chapter 3

Methodology

This section describes that the development plan of Breast Cancer Detector Software-CBR will be using the method development Life Cycle (SDLC).

3.1 Introduction.

The purpose of this study is to develop a software that will be use by doctor who expertise in cancer breast. This software will make doctor easier to find a case based on old files that have the same cases as the new patient. This will shows that patient on what stage of cancer based on old patient record. In the process of develop a software, we need some methodology to make the software more better and create like a planned. Methodology is about the planning the process to develop the software product. Methodolgy can be used to know the cost will be used, hardware and software will be used, achieve goal in time and so on. The purpose of this chapter are to describe the flow of research that will be make.

The data from past patient or old record can be useful to find level of cancer new patient by calculation of case based reasoning.

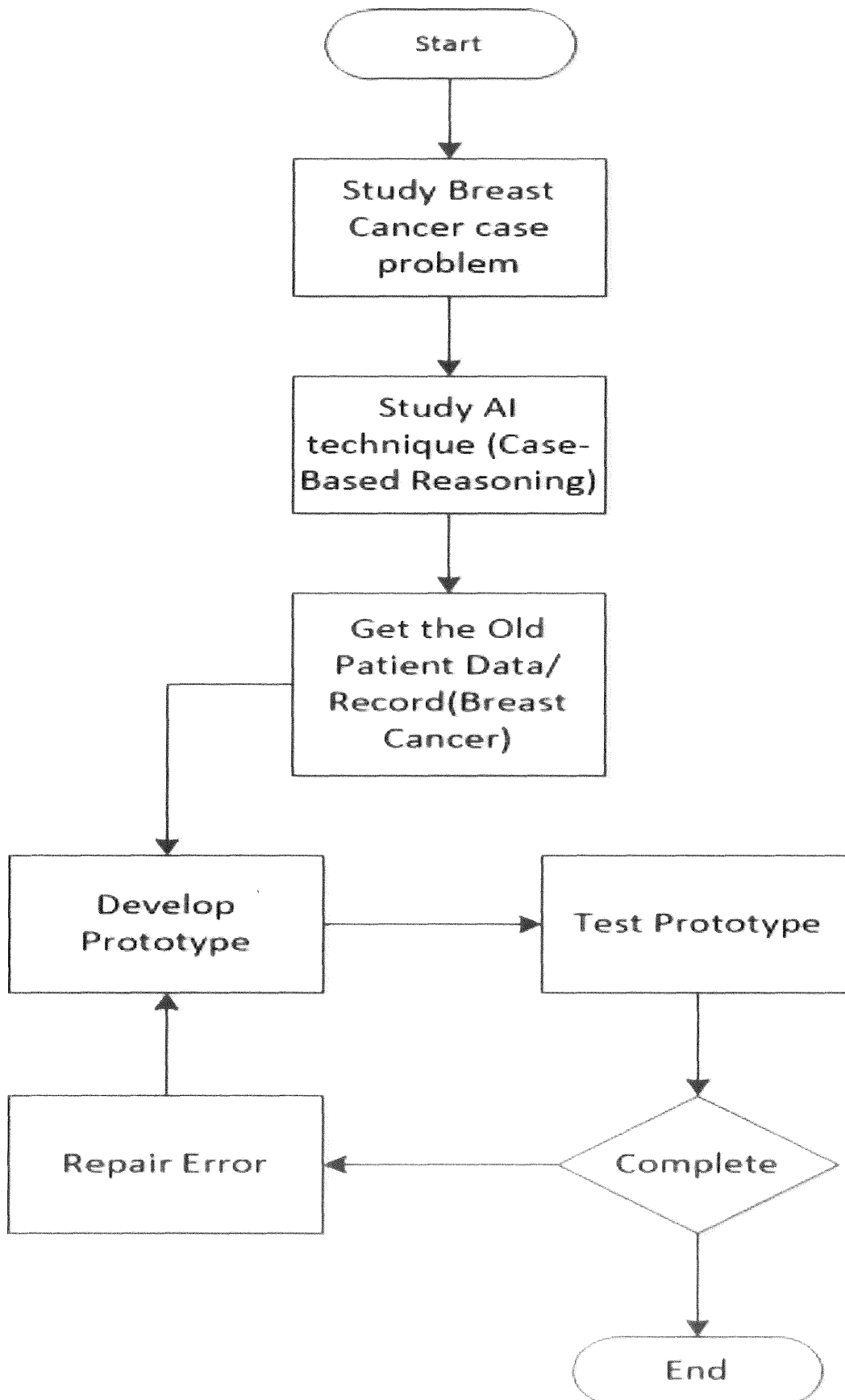
Flowchart of the research

Figure 2: Flowchart

3.2 Methodology.

For the software development methodologies, I use the custom methodology because to make sure my project to be plan efficiently.

3.2.1 Proposed Methodology

3.2.1.1 Steps Involves In Methodology(Analysis)

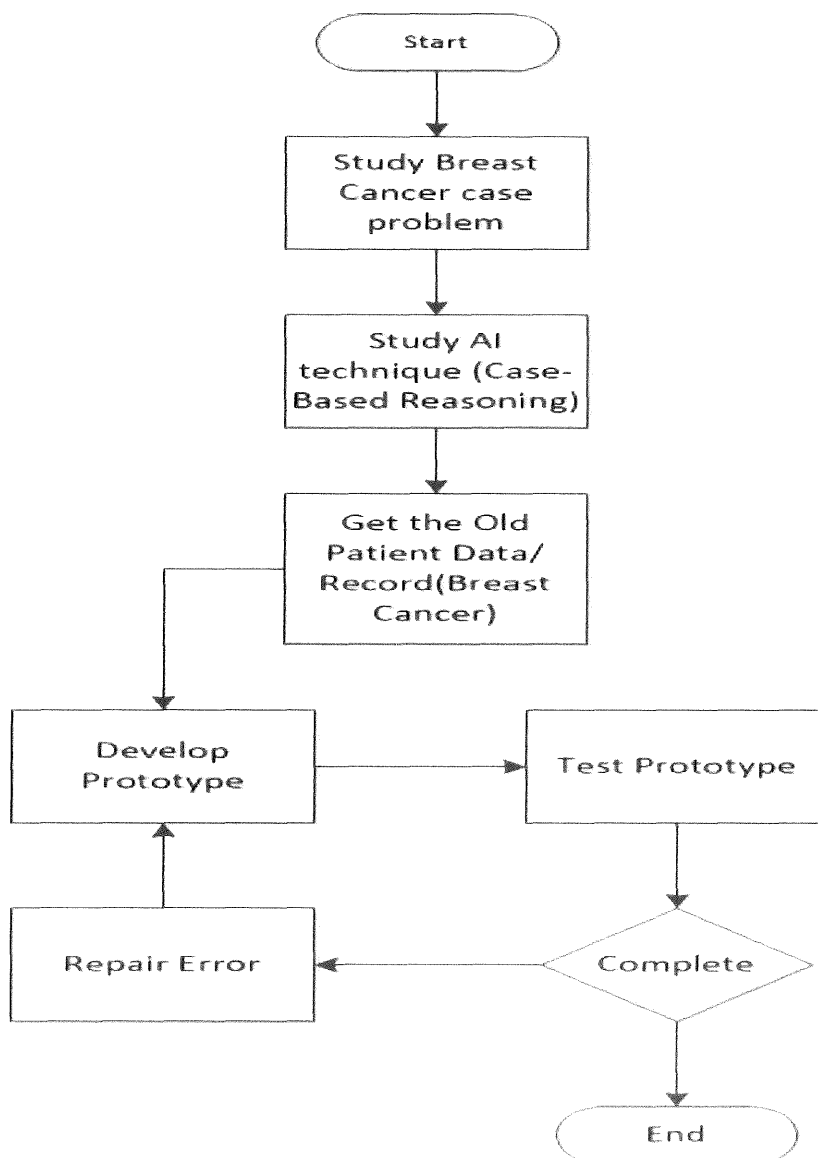


Figure 3: Phase of Methodology

Phase	Activity
Study Breast Cancer Case Problem	Study the breast cancer attribute, level, type cancer and create problem statement.
Study AI Technique (Case-Based Reasoning)	Study Artificial Intelligent technique which is the process of case based reasoning calculation. The calculation is called feature based similarity measure.
Get The Old Patient Data / Record (Breast Cancer)	Find the record (dataset) patients that already resolved their case.
Maintenance	Maintenance the software prototype : <ol style="list-style-type: none"> 1) Develop prototype 2) Test prototype 3) Repair error prototype

Table 3: Describe Methodology

3.2.1.2 Proposed Technique

Case-Based Reasoning (CBR) are one of the technique artificial intelligent. This technique is call as artificial intelligent is because of their action is like a human thinking. The comparable technique is just like the human based on the calculation. CBR works by compare the new data and existing data to find the result of the case. CBR use the similarity measured calculation to find the result of the new data.

There are four similarity measure calculation :

5. Distance-Based Similarity Measure
6. Probability Similarity Measure
7. Feature-Based Similarity Measure
8. Classification Similarity Measure

The CBR of my study is use the Feature-Based similarity measure calculation.

3.2.2 Hardware & Software

Hardware & Software	Description	Importance
1) Asus A43SV Series 2) ACER Aspire 4738G	Used to Saving all the project and Software to be used	All information are stored here
IDM 6.15	To download information Online	Faster time compare to other download manager.
Netbeans IDE 8.0.2	Software that will be used in developing which consist a java language	Main software for programming.
Microsoft Word 2007/2010	Software used to make documentation.	Make it easier with many tools for making document.
Microsoft Project 2007/2010	Software use to make Gantt chart	Make schedule more effective and precise.
Microsoft Visio 2007/2010	Use to create flowchart, data flow and work flow	Creating shape and arrow of the data flow easier.

Table 4: List of Software and Hardware

3.3 Gantt Chart.

Show in section **APPENDIX A.**

4.3 Constraints

Constraints for this project are categorized into two parts:

4.3.1 Development constraints

4.3.2 System constraints

4.3.1 Development constraints

There many data set and data will be updated every day. In 2010, nearly 1.5 million people get breast cancer. New data should be updated every day and should be collect that size of worldwide.

4.3.2 System constraints.

System should interact to all of the hospital in the worldwide to get new data everyday or month which is used high cost to develop the system all over the world.

4.4 Developing the Software.

In netbeans IDE, programmer can create the design of the interface using the existing items build in the netbeans. Programmer can drag and drop the item which is the button, labels, text, list or else. After dropping the items, netbeans automatically generate the code in the script. Programmer just add little more coding to create the function.

Project that have been create can be select here to see the script or interface.

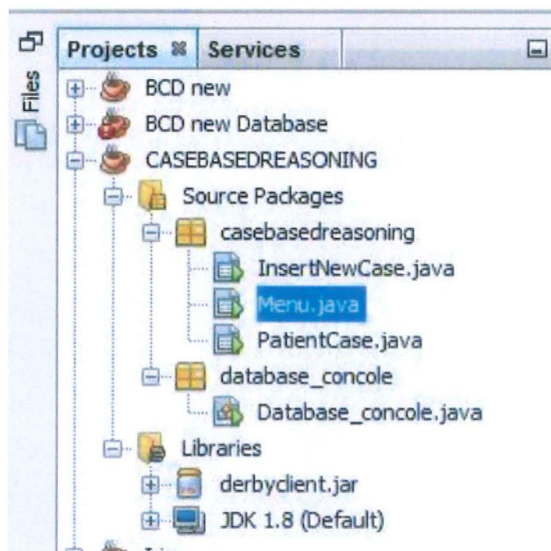


Figure 4: Editor Toolbar project script

Interface editor java is called JFrame.java. As we can see, at the right corner is the tool that programmer can drag and drop at the JFrame.java. At the bottom right corner is the properties that dragged items to edit the size, font, names items (example: lbl for label, btn for button) and so on.

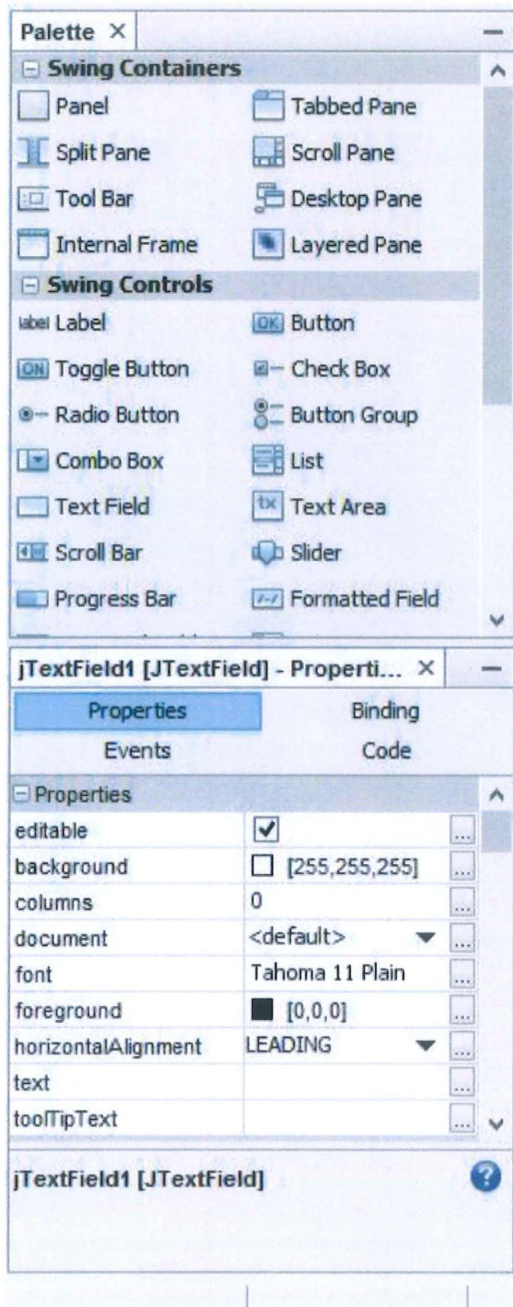


Figure 5: JFrame.java (palette and properties)

4.5 Result of the Application

After successful build the software from netbeans, the output is the case that similarity to another case. Software will show the class or level of the cancer which is 2 for benign and 4 for malignant. If new case are not similar to existing case (data set), software will show the percentage of the similarity.

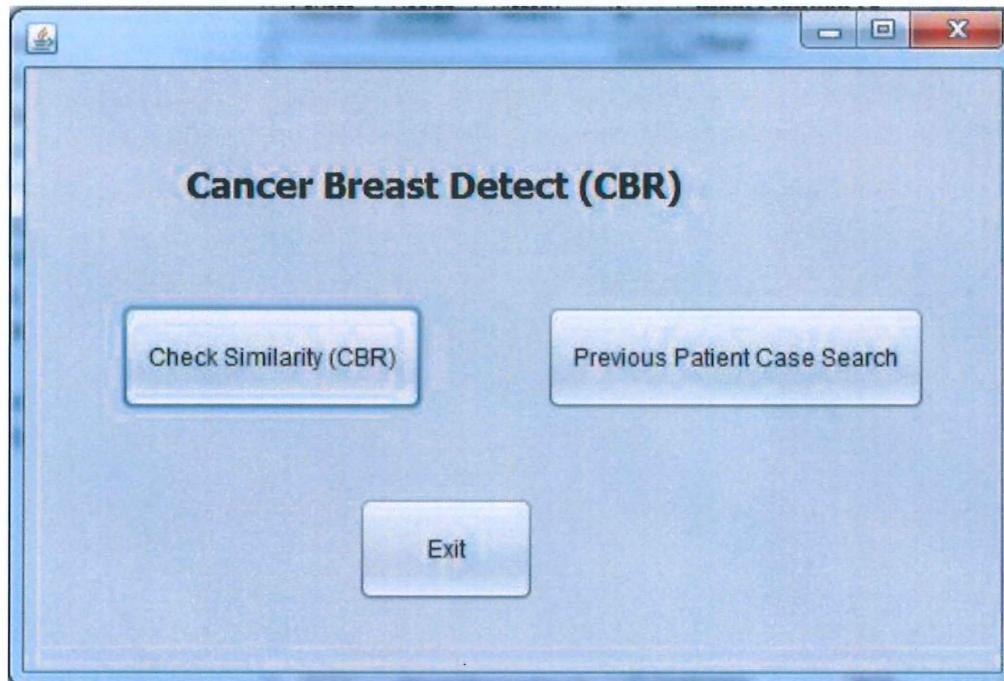


Figure 6: Menu.java (Menu)

CHECK SIMILARITY (CBR CALCULATION)

Clump Thickness: (1 - 10)	<input type="text" value="5"/>	Result : Similarity at 83.95% Highest similarity : Highest similar to case number : 627 Sample Code : Sample Code: 1190386 Case Level : Malignant
Uniformity of Cell Size: (1 - 10)	<input type="text" value="5"/>	
Uniformity of Cell Shape: (1 - 10)	<input type="text" value="5"/>	
Marginal Adhesion: (1 - 10)	<input type="text" value="5"/>	
Single Epithelial Cell Size: (1 - 10)	<input type="text" value="5"/>	
Bare Nuclei: (1 - 10)	<input type="text" value="5"/>	
Bland Chromatin: (1 - 10)	<input type="text" value="5"/>	
Normal Nucleoli: (1 - 10)	<input type="text" value="5"/>	
Mitoses: (1 - 10)	<input type="text" value="5"/>	

Figure 7: InsertNewCase.java (Insert New Case and Calculate (CBR))

The screenshot shows a Java application window titled "Patient Case". The window has a blue background and a standard Windows-style title bar with minimize, maximize, and close buttons. The main content area is light blue and contains the following elements:

- Header:** "Patient Case" in bold black text.
- Search Section:** "Patient Number : 1002945" with a text input field containing "1002945". To the right is a "Find Case" button.
- Result Section:** A white rectangular box containing the following text:

```
NUMBER : 2
Patient Number (Sample Code) : 1002945
Clump Thickness : 5
Unified Cell Size : 4
Unified Cell Shape : 4
Marginal Adhesion : 5
Single Epithelial Cell Size : 7
Bare Nuclei : 10
Bland Chromatin : 3
Normal Nucleoli : 2
Mitoses : 1
Class : Benign
```
- Navigation:** A "Back" button is located to the right of the result box.

Figure 8: PatientCase.java (Search Previous Patient Record by SAMPLE CODE)

The dataset was stored in the database that is use the Derby.jar database.

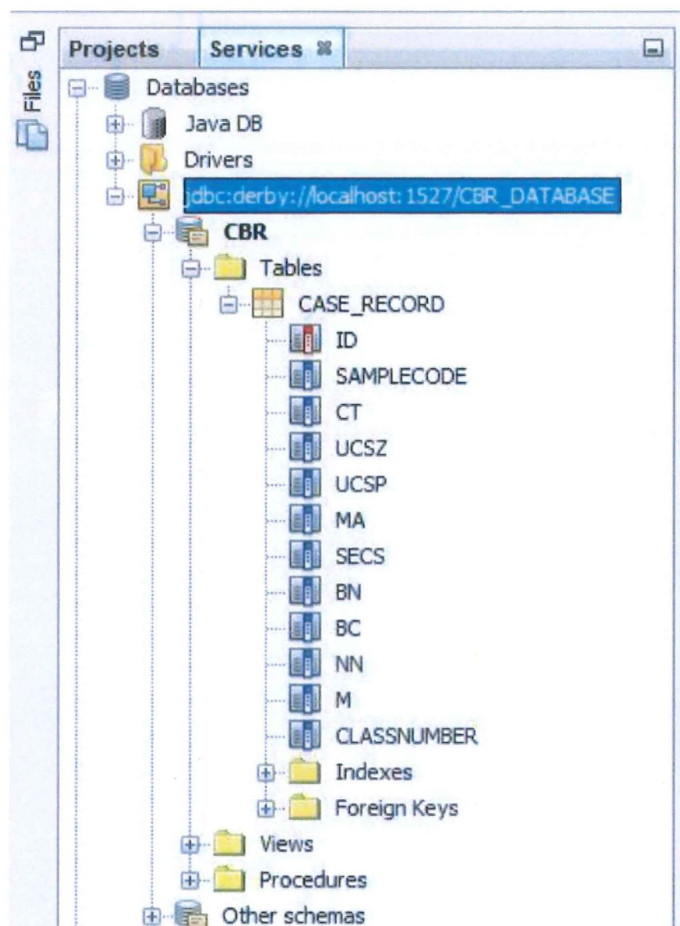


Figure 9: Database (use driver Derby.jar)

4.6 Coding and Explanation

4.6.1 Coding CBR – Featured Based Reasoning

This coding is build based on the logic CBR calculation technique. Equation of the Feature Based Reasoning is interprete into java coding.

Code:

```
public void calculate() {

    double Result_CT, Result_UCSZ, _UCSP, Result_MA, Result_SECS, Result_BN,
    Result_BC, Result_NN, Result_M, First_formula, Second_formula, Third_formula,
    Result_Total;

    int caseNumberShow=0;

    int cn = 0;

    for (int i = 0; i < 699; i++) {

        //local similarity

        caseNumberShow = SAMPLECODE[i];

        cn = CLASSNUMBER[i];

        if(cn == 2){

            ClassType = "Benign";    }

        else if(cn == 4){

            ClassType = "Malignant"; }

        Result_CT = (double) Math.abs((getCT - CT[i]) / 9);

        Result_UCSZ = (double) Math.abs((getUCSZ - UCSZ[i]) / 9) ;
```

```
Result_UCSP = (double) Math.abs((getUCSP - UCSP[i]) / 9);
```

```
Result_MA = (double) Math.abs((getMA - MA[i]) / 9);
```

```
Result_SECS = (double) Math.abs((getSECS - SECS[i]) / 9);
```

```
Result_BN = (double) Math.abs((getBN - BN[i]) / 9);
```

```
Result_BC = (double) Math.abs((getBC - BC[i]) / 9);
```

```
Result_NN = (double) Math.abs((getNN - NN[i]) / 9);
```

```
Result_M = (double) Math.abs((getM - M[i]) / 9);
```

```
First_formula = Math.abs (((2 * Result_CT) + (1 * Result_UCSZ) + (1 *
Result_UCSP) + (1 * Result_MA) + (1 * Result_SECS) + (1 * Result_BN) + (1 *
Result_BC) + (1 * Result_NN) + (1 * Result_M)));
```

```
Second_formula = First_formula / 9;
```

```
Third_formula = 1 - Second_formula;
```

```
Result_Total = Third_formula * 100;
```

```
if(Result_Total > similarityResult) //Contender greater than current highest
similarity
```

```
{
```

```
    similarityResult = roundUp(Result_Total); //new highest result
```

```
    similarityLocation = i;
```

```
}
```

```
}
```

```
sampleCodeDisplay = caseNumberShow;
```

```
}
```

```
double roundUp(double i_input) //Rounding up function
```

```
{
```

```
    DecimalFormat twoDForm = new DecimalFormat("#.##");
```

```
    return Double.valueOf(twoDForm.format(i_input));
```

```
}
```

4.6.2 Read Data Set from Database

This coding is to read the data set from text file (Derby.jar) line by line.

Code:

```
try{

    String host = "jdbc:derby://localhost:1527/CBR_DATABASE";

    String uName = "CBR";

    String uPass= "CBR123";

    Connection con = DriverManager.getConnection(host, uName, uPass);

    Statement stmt = con.createStatement( );

    String sql = "SELECT * FROM CASE_RECORD";

    ResultSet rs = stmt.executeQuery(sql);

    loop = 0;

    while(rs.next()){

        int id1 = rs.getInt("ID");

        int SAMPLECODE1 = rs.getInt("SAMPLECODE");

        int CT1 = rs.getInt("CT");

        int UCSZ1 = rs.getInt("UCSZ");

        int UCSP1 = rs.getInt("UCSP");
```

```
int MA1 = rs.getInt("MA");

int SECS1 = rs.getInt("SECS");

int BN1 = rs.getInt("BN");

int BC1 = rs.getInt("BC");

int NN1 = rs.getInt("NN");

int M1 = rs.getInt("M");

int CLASSNUMBER1 = rs.getInt("CLASSNUMBER");

ID[loop] = id1;

SAMPLECODE[loop] = SAMPLECODE1;

CT[loop] = CT1;

UCSZ[loop] = UCSZ1;

UCSP[loop] = UCSP1;

MA[loop] = MA1;

SECS[loop] = SECS1;

BN[loop] = BN1;

BC[loop] = BC1;

NN[loop] = NN1;

M[loop] = M1;

CLASSNUMBER[loop] = CLASSNUMBER1;

loop = loop + 1;

}          //while loop end
```

```
}  
  
catch (SQLException err){  
    System.out.println( err.getMessage() );  
}
```

4.7 System Limitation

This application is build only to show the CBR technique and this software is only have one function. This is the simple software to show how it work and it working like a human. This software is work like a logic thinking human to solve the problem by comparing the situation.

4.8 Advantage of this Application

There are five advantage of this software:

- 1) Reduces the knowledge accusation effort – Knowledge-Based System
 - Correctness of the rule-base developed can be questionable due to the knowledge gap between knowledge engineer and domain expert.

- 2) White-box learning – explanation facilities
 - CBR stores old experiences or cases in a memory known as a case - base. This enables the algorithm to recall previous cases to solve a new problem. This process is called retrieval.

- 3) Incremental learning
 - an ability of a learning algorithm to incorporate new knowledge into existing knowledge to suite the available data. Learning new data without forgetting prior knowledge is an issue in learning algorithm

- 4) The closest match to human reasoning
 - Closer to human thinking and solving problem. Use case base reasoning is a technique or the way people think. Identifying the problem through experiences that have passed and if no experience, it will be a new experience for future use.

- 5) Makes use of existing data, (example: database)
 - Waste of old records of cancer patients do not occur. Old records can be used to compare the similarity of new record the patient's condition critical level for new patients.

- 6) Saving time and cost
 - Time and cost to do all the test for new case and it happened like the old case that have been solve.

4.9 Future Enhancement

In the future, some function will be repair or add in this software

- 1) Bigger database that will save all the possible data set.
- 2) Function that expertise doctor can add the new cases from the internet or portal
- 3) Widely use of this technique and software at hospital or clinic all over the world.

Chapter 5

Conclusions

5.1 Introduction

This application will make the user (doctor) easily to compare new breast cancer case to previous case which is have many record (dataset). This application based on feature-based similarity measure calculation which is the algorithm to check the similarity of the new attribute with the previous attribute like new breast cancer case and previous breast cancer case. In the Breast Cancer Detector software have a function to check similarity where is I take from algorithm of case-based reasoning technique and turn it into a coding. Thus, in the future, which is during PSM II, the courseware application is expected to be developed as how it has been planned and it is highly hoped that the objectives and the goals of the courseware will be achieved at the end.

5.2 Summary of Literature Review

In literature review, developer will know more details about problem statement, scope ad current application comparison as this information could help developer to design and develop the application. More information is collected that related to this application. Based on the problems faced by currently available systems that will motivate the developer to develop this project.

5.3 Summary of Methodoly

In the course of this undertaking, all believed exactly how critical this undertaking which is making a Breast Cancer Detector (CBR). Inside undertaking technique, the process helpful to develop this undertaking is identified a protracted using the objective of while using this method. . The particulare undertaking demands make clear much more concerning info get together, software as well as computer hardware of which required to develop this undertaking, in which there components are crucial specifically

inside developing model of the properties and environment which often to have a larger specs regarding personal computer.

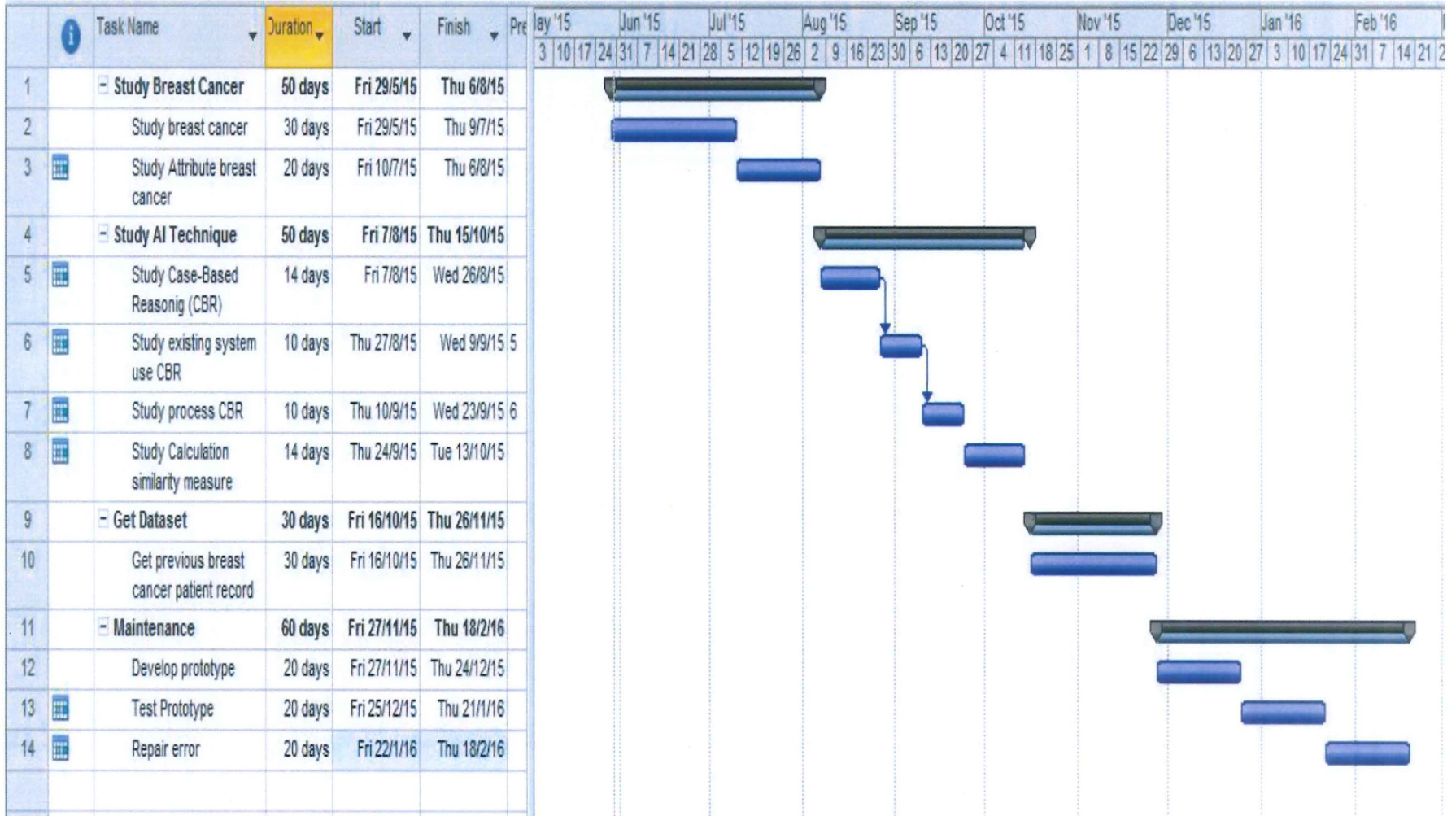
References

- [1] Dr. William H. Wolberg, Olvi Mangasarian (1990). *General format*. Retrieved from [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Original\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original))
- [2] Zhang, J. (1992), *General format*. Retrieved from <http://www.cancer.gov/types/breast>
- [3] Sujana Movva, MD on (March 15, 2015). *General format*. Retrieved from <http://www.webmd.com/breast-cancer/>
- [4] Agnar Aamodt, Enric Plaza (1994). *General format*. Retrieved from <http://www.iiia.csic.es/~enric/papers/AICom.pdf>
- [5] Janet L. Kolodner (1992). *General format*. Retrieved from http://alumni.media.mit.edu/~jorkin/generals/papers/Kolodner_case_based_reasoning.pdf
- [6] W. Mark, E. Simoudis and D. Hinkle, “Case-based reasoning: Expectations and results”, *Case-Based Reasoning: Experiences, Lessons & Future Directions*, D.B. Leake, (Ed.), pp. 269-294, AAAI Press, Menlo Park, USA, 1996.
- [7] J. Kolodner, “Making the implicit explicit: Clarifying the principles of case-based reasoning”, *Case-Based Reasoning: Experiences, Lessons & Future Directions*, D.B. Leake, (Ed.), pp. 349-370, AAAI Press, Menlo Park, USA, 1996.
- [8] M. Pantic and L.J.M. Rothkrantz, “Case-based reasoning for user-profiled recognition of emotions from face images”, *Proc. IEEE Int’l Conf. Multimedia and Expo*, 2004.
- [9] I. Watson and F. Marir, “Case-base reasoning: A review”, *The Knowledge Engineering Review*, vol. 9, no. 4, pp. 327-354, 1994.
- [10] Petite Hierro Roberto, (1997). *General format*. Retrieve from <http://exit.udg.edu/people.aspx#assistants>
- [11] K Amailef, “J Lu –Decision Support Systems”, (2013). *General format*. Retrieve from <http://www.sciencedirect.com/science/article/pii/S0167923613000043>

[12] M Minor, S Montani, Luc Lamontagne and Juan A. Recio-Garcia,” Process Case-based Reasoning”, Process-oriented Case-based Reasoning Review, 2012.

APPENDIX

APPENDIX A : PROJECT GANTT CHART





Breast Cancer Detector (CBR)

Prepared by: Ahmad Bukhari Abdullah

Supervised by : Azlina Zainuddin

UNIVERSITY MALAYSIA PAHANG

Introduction

Project that build to detect breast cancer patients level use the Artificial Intelligence technique that is called Case-Based Reasoning calculation by make use the old dataset of the record cases. This system will show the similarity measure percentage from all of the set old data. A case-based reasoned solves new problem by using or adapting solution that were used to solve old problems. A case-based reasoning is an AI technique that imitates how human make a decision. In CBR, new problems are solved by recalling from a previously solved problem which are stored in the case-base.

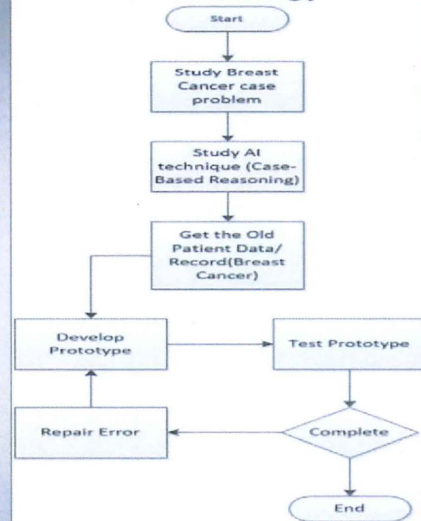
Objective

- 1) To study case based reasoning artificial intelligent algorithm to apply in breast cancer detector software.
- 2) To de3velop a tool to process the similarity measurement and make the compares from the new record with previous record of the patient.
- 3) To evaluate the tool for improvement of detection of the cancer patients history.

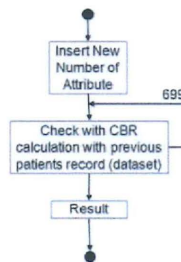
Scope

- 1) This project concern on patients who related to the cancer detection focused on breast cancer.
- 2) This project also concern on doctor specialized in breast cancer.
- 3) The developed tool will use CBR calculation only.
- 4) Record of patient will use from Dr. William H. Wolberg (physician) from University of Wisconsin Hospitals, Madison, Wisconsin, USA dataset. ([https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Original\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original)))

Methodology



Flowchart



Results

Patient Case

Patient Number :

NUMBER : 2
 Patient Number : 5
 Unified Cell Size : 4
 Unified Cell Shape : 4
 Marginal Adhesion : 5
 Single Epithelial Cell Size : 7
 Bare Nuclei : 10
 Bland Chromatin : 3
 Normal Nucleoli : 2
 Mitoses : 1
 Class : Benign

NEW CASE

Clump Thickness (1 - 10) :

Uniformity of Cell Size (1 - 10) :

Uniformity of Cell Shape (1 - 10) :

Marginal Adhesion (1 - 10) :

Single Epithelial Cell Size (1 - 10) :

Bare Nuclei (1 - 10) :

Bland Chromatin (1 - 10) :

Normal Nucleoli (1 - 10) :

Mitoses (1 - 10) :

Conclusions

In the future, some function I wish to be repair or add in this software. First is bigger database that will save all the possible data set. Second is that expertise doctor can add the new cases from the internet or portal. Third is widely use of this technique and software at hospital or clinic all over the world.