BENCHMARKING METHODOLOGY FOR MULTICLASS CLASSIFICATION MODELS BASED ON MULTI CRITERIA DECISION ANALYSIS

MOHAMMED ASSIM MOHAMMED ALI

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

SULTAN IDRIS EDUCATION UNIVERSITY

2019

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

BENCHMARKING METHODOLOGY FOR MULTICLASS CLASSIFICATION MODELS BASED ON MULTI CRITERIA DECISION ANALYSIS

MOHAMMED ASSIM MOHAMMED ALI

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENT FORTHE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF ART, COMPUTING & CREATIVE INDUSTRY
SULTAN IDRIS EDUCATION UNIVERSITY

2019

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

UPSI/IPS-3/BO 32
Pind : 00 m/s: 1/1

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

Please tick (✓)

Project Paper	
Masters by Research	
Masters by Mix Mode	
Ph.D.	✓

INSTITUTE OF GRADUATE STUDIES

DECLARATION OF ORIGINAL WORK

This declaration is made on the 18 day of 06 / 2019

i.Student's Declaration:
I'm Mohammed Assim Mohammed Ali Alsalem - P20161001045-Faculty of Art, Computing, and Creative Industry Hereby declares that the dissertation / thesis for titled (Benchmarking methodology for multiclass classification models based on multi criteria decision analysis) is my original work. I have not copied from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has part been written for me by another person.

_____
Signature of the student

ii. Supervisor's Declaration:
I'm Dr. Mashitoh Hashim hereby certify that the work entitled (benchmarking methodology for multiclass classification models based on multi criteria decision analysis) was prepared by the above-named student, and was submitted to the Institute of Graduate Studies as a partial / full fulfillment for the conferment of the requirements for Doctor of Philosophy (By Research), and the aforementioned work, to the best of my knowledge, is the said student's work.

_____ _____

Date Signature of the Supervisor

UPSI/IPS-3/BO 31
Pind.: 01 m/s:1/1

INSTITUT PENGAJIAN SISWAZAH /
INSTITUTE OF GRADUATE STUDIES

BORANG PENGESAHAN PENYERAHAN TESIS/DISERTASI/LAPORAN KERTAS PROJEK
DECLARATION OF THESIS/DISSERTATION/PROJECT PAPER FORM

Tajuk / *Title*: Benchmarking methodology for multiclass classification models based on multi criteria decision analysis

No. Matrik /*Matric No.*: P20161001045

Saya / *I* : Mohammed Assim Mohammed Ali Alsalem

mengaku membenarkan Tesis/Disertasi/Laporan Kertas Projek (Kedoktoran/Sarjana)* ini disimpan di Universiti Pendidikan Sultan Idris (Perpustakaan Tuanku Bainun) dengan syarat-syarat kegunaan seperti berikut:-
acknowledged that Universiti Pendidikan Sultan Idris (Tuanku Bainun Library) reserves the right as follows:-

1. Tesis/Disertasi/Laporan Kertas Projek ini adalah hak milik UPSI.
   The thesis is the property of Universiti Pendidikan Sultan Idris

2. Perpustakaan Tuanku Bainun dibenarkan membuat salinan untuk tujuan rujukan dan penyelidikan.
   Tuanku Bainun Library has the right to make copies for the purpose of reference and research.

3. Perpustakaan dibenarkan membuat salinan Tesis/Disertasi ini sebagai bahan pertukaran antara Institusi Pengajian Tinggi.
   The Library has the right to make copies of the thesis for academic exchange.

4. Sila tandakan (√) bagi pilihan kategori di bawah / Please tick (√) from the categories below:-

☐ **SULIT/*CONFIDENTIAL*** — Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub dalam Akta Rahsia Rasmi 1972. / *Contains confidential information under the Official Secret Act 1972*

☐ **TERHAD/*RESTRICTED*** — Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan ini dijalankan. / *Contains restricted information as specified by the organization where research was done.*

☐ **TIDAK TERHAD / *OPEN ACCESS***

(Tandatangan Penyelia / *Signature of Supervisor*)

(Tandatangan Pelajar/ Signature)

& (Nama & Cop Rasmi / *Name & Official Stamp*)

Tarikh: _____

Catatan: Jika Tesis/Disertasi ini **SULIT** @ **TERHAD**, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh laporan ini perlu dikelaskan sebagai **SULIT** dan **TERHAD**.

Notes: If the thesis is CONFIDENTAL or RESTRICTED, please attach with the letter from the related authority/organization mentioning the period of confidentiality and reasons for the said confidentiality or restriction.

ACKNOWLEDGMENT

"In the name of Allah, the Most Gracious and the Most Merciful"

Alhamdulillah, first and foremost, praise be Allah, the Cherisher and Sustainer of the World and to the Prophet Muhammad (Peace and Blessings of Allah Be Upon Him) who was sent by Allah to be a great teacher to the mankind. I would like to extend my appreciation to those who involved and give a helpful hand in ensuring the success of this research. This research would not have come to fruition without all your help and supports. I am very grateful to my supervisors Dr.Mashitoh Hashim and Dr. Aos Alaa Zaidan for guiding me during my work on this research. I would like to express my sincere thanks and gratitude to them for their continuous guidance, support and patience. I would like to express my sincere thanks to Dr. Bilal Bahaa Zaidan for help me and support as well, my appreciation and gratitude for him. I am also very grateful to my family especially my father and mother for their support blessing, patience, love, and encouragement since my childhood. I extend my love and appreciation to my wife Noor for her continuous support and encouragement. Finally, I would like to thank all friends who have helped me and encouraged me. Thank you. Allah blesses you.

ABSTRACT

The purpose of this research was to develop a benchmarking methodology for aiding medical organizations' administrations in benchmarking and ranking available multiclass classification models to select the best one. Medical organizations have been facing difficulties in evaluating and comparing classification models. Experimental and case study research methods were adopted in this study. The new benchmarking methodology was proposed based on two stages. In first stage, a Decision Matrix (DM) was constructed based on the crossover of two groups of multi-evaluation criteria and 22 multiclass classification models. The matrix was evaluated using secondary datasets consisting of 72 samples of acute leukemia, including 5327 gens. In the second stage, multi-criteria decision-making techniques, namely, Best and Worst method (BWM) and Vlse Kriterijumska Optimizacija Kompromisno Resenje (VIKOR) were used to benchmark and ranked the multiclass classification models. The BWM was applied to calculate the weights of evaluation criteria, whereas VIKOR was used to benchmark and rank the multi-class classification models. VIKOR was utilized in two decision-making contexts, namely individual and group contexts. In group decision making, internal and external group aggregations are applied. For validating the proposed methodology, an objective method was used. The results showed that (1) the integration of BWM and VIKOR was effective for solving the benchmarking/selection problems of multi-class classification models. (2) The ranks of multi-class classification models obtained from internal and external VIKOR group decision making were almost the same, where, Bayes. Naïve Byes Updateable, Bayes Net, Decision Stump were the first three classification models respectively and Trees. LMT was the last one. (3) In the objective validation, the ranking results of internal and external VIKOR group decision making were valid. Clearly, as a conclusion, the proposed methodology can be used for evaluation and benchmarking different multiclass classification models for various applications. The implications of this study will benefit medical organizations by enabling them to make the right decisions regarding the use of multi-class classification models for acute leukemia and the implications also benefit medical classification software developers who work in industrial companies and institutions in developing classification models.

METODOLOGI PENANDA ARAS UNTUK MODEL PENGELASAN MULTICLASS BERDASARKAN ANALISIS PELBAGAI KRITERIA

ABSTRAK

Penyelidikan ini bertujuan untuk membangunkan satu metodologi penanda aras bagi membantu pentadbiran organisasi perubatan dalam penanda-arasan dan penentuan kedudukan model klasifikasi pelbagai kelas yang sedia ada untuk memilih model yang terbaik. Buat masa ini, organisasi perubatan menghadapi kesukaran dalam menilai dan membandingkan model klasifikasi. Kaedah kajian eksperimen dan kajian kes digunakan dalam kajian ini. Oleh itu, satu matriks keputusan dicadanglcan berdasarkan dua fasa. Dalam fasa pertama, satu matriks keputusan dibangunkan berdasarkan lintasan dua kumpulan kriteria pelbagai penilaian dan 22 model klasifikasi pelbagai kelas. Matriks berkenaan dinilai menggunakan satu dataset sckunder yang melibatkan satu sampel terdiri daripada 72 pesakit leukemia akut, termasuk 5327 gen. Dalam fasa kedua, teknik pembuatan keputusan pelbagai kriteria, iaitu Best-and-Worst Method (BWM) dan Vlse Kriterijumska Optimizacija Kompromisno Resenje (VIKOR), digunakan untuk menanda aras dan menentukan kedudukan model-model klasifikasi pelbagai kelas. Lebih terperinci lagi, BWM digunakan untuk mengira pemberatan kriteria penilaian, manakala VIKOR digunakan untuk menanda aras dan menentukan kedudukan model klasifikasi, di mana teknik VIKOR menggunakan pcngagregatan kumpulan dalaman dan luaran. Kaedah objektif digunakan untuk menilai metodologi yang dicadangkan. Dapatan menunjukkan (1) pengintegrasian BWM dan VIKOR adalah berkesan dalam menyelesaikan masalah penanda-arasan dan pemilihan model klasifikasi pelbagai kelas. Tambahan pula, (2) kedudukan klasifikasi model yang diperoleh dari keputusan kumpulan dalaman dan luaran VIKOR adalah hampir sama, Bayes. Naïve Byes Updateable, Bayes Net, Decision Stump adalah tiga model klasifikasi pertama dan Trees. LMT adalah yang terakhir. (3) Selanjutnya, dapatan kedudukan yang diperoleh dari keputusan VIKOR telah dibuktikan kesahannya. Maka, implikasi dari pengunaan metodologi penanda-arasan ini adalah organisasi perubatan dapat membuat keputusan yang tepat mengenai penggunaan klasifikasi model untuk pesakit leukemia akut dan pembangun perisian klasifikasi perubatan dapat membangun beberapa model yang berbeza untuk pelbagai aplikasi dengan lebih berkesan lagi. Sebagai kesimpulan, metodologi yang dicadangkan boleh digunakan untuk penilaian dan penandaarasan model klasifikasi multiclass yang berbeza untuk pelbagai aplikasi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DMs	Decision makers
GDM	Group Decision Making
DM	Decision Matrix
MCDM	Multi- Criteria Decision Making
MCDA	Multi-Criteria Decision Analysis
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
C	Criteria
FP	False Positive
FN	False Negative
TP	True Positive
TN	True Negative
MOP	Matrix of parameters
ROP	Relationship of Parameters
BOP	Behavior of Parameters
ER	Error Rate
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
VIKOR	Vlse Kriterijumska Optimizacija Kompromisno Resenje
MEW	Multiplicative Exponential Weighting
SAW	Simple Additive Weighting
WSM	Weighted Sum Mode
AHP	Analytic Hierarchy Process
ANP	Analytic Network Process
BWM	Best-Worst-method
CR	Consistency Ratio
ML	Machine Learning
ANN	Artificial Neural Network

SVM	Support Vector Machine
WEKA	Waikato Environment for Knowledge Analysis
SD	Standard deviation

APPENDICES LIST

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter present a brief background about the research, the state of the problem, the motivation of this research, and the research objectives. In Section 1.2, a brief background about the research components is presented. In Section 1.3, the research problem. In Section 1.4 and Section 1.5, research questions and research objectives are reported respectively, as well as in section 1.6, relationship between research objectives, research questions and research problem. In section 1.7, the scope of the research. In Section 1.8, the motivation. In Section 1.9, significance of the study. Section 1.10 presents the main terms of this study. Finally, outline the main structure of the thesis are briefly reported in section 1.11.

1.2 Background of study

The administration of medical organizations must make the right decisions, particularly in selecting automated solutions related to the diagnosis and detection of complex diseases, such as acute leukemia. The importance of this decision is due to the widespread use (Srisukkham, Zhang, Neoh, Todryk, & Lim, 2017) and the actual need to use them (Agaian, Madhukar, & Chronopoulos, 2014). Many researchers, such as (Agaian et al., 2014; Bagasjvara, Candradewi, Hartati, & Harjoko, 2016; Labati, Piuri, Scotti, & Ieee, 2011; Lei & Chen, 2012; S. Mohapatra, Patra, & Satpathi, 2010; Jyoti Rawat, Singh, Bhadauria, & Virmani, 2015; Srisukkham et al., 2017) have confirmed that automated solution based on artificial intelligence techniques can provide rapid acute leukemia diagnosis and classification as well as increase the reliability and accuracy of results. Furthermore, several physicians, cancer treatment centers and hospitals have started using automated models for acute leukemia classification to face the several potential limitations of manual analysis (Agaian et al., 2014; Mohapatra et al., 2010; Srisukkham et al., 2017).

However, the growing numbers of those automated classification models have become more challenging for the users who are looking for models that deliver high accurate results in a short time with no-errors(Snousy, El-Deeb, Badran, & Khlil, 2011). Therefore, the administrations of health organizations have been facing difficulties in evaluating and benchmarking automated classification models for the classification of acute leukemia to determine the best model, especially because no single model is superior to the rest (Agaian et al., 2014; Goutam & Sailaja, 2015;

Snousy et al., 2011) Furthermore, many of these models suffer from lack of accuracy and computational efficiency (Mishra, Majhi, Sa, & Sharma, 2017) and evaluation and comparison become complicated due to the presence of multiple evaluation criteria (Nguyen & Nahavandi, 2016). The evaluation and benchmarking of automated classification tasks for serious medical cases, such as acute leukemia, are crucial in obtaining the best result.

The evaluation and benchmarking processes of automated classification tasks for serious medical cases such as acute leukemia is crucial in the quest for getting the best result (Labati et al., 2011). For example, if the model incorrectly identifies non-cancer cells as cancerous, then this error may result in adverse effects on the patient's mental state, and the patient will need further diagnosis and surgery to determine whether he is cancer-free. The most serious case is when the model incorrectly identifies cancer cells as noncancerous. On this basis, the most efficient technique must be sought to help medical organizations select the suitable classification model for acute leukemia. Therefore, evaluation and benchmarking processes for determining the best automated multiclass classification model among many available alternatives are required, especially because these models are expensive and related to the medical aspect of humans (Mishra et al., 2017).

The procedures of evaluation and benchmarking the automated model for multiclass classification of acute leukemia is a challenging problem (Agaian et al., 2014). Evaluation and benchmarking in order to enable health organizations to choose

the automated model for acute leukemia that provides the best result is a difficult decision-making task with several measurements (Goutam & Sailaja, 2015).

Many studies have discussed the development of automated models for acute leukemia analysis, how to use the models and the benefits that health organizations can gain from using them (Agaian et al., 2014; Bhattacharjee & Saini, 2015; Goutam & Sailaja, 2015; Nazlibilek et al., 2014; Jyoti Rawat et al., 2015; Singhal & Singh, 2014; Torkaman, et.al, 2009; Wang & Palade, 2007); However, few studies have aimed to assist health organizations in evaluating and benchmarking among the available classification models and to determine the best model. Current literature on the evaluation and benchmarking of automated multiclass classification models for acute leukemia is limited and scattered. A few studies have discussed the evaluation and benchmarking of automated classification for acute leukemia, but they are limited to one aspect of performance evaluation while neglecting other aspects (Cornet et al., 2008; Krappe et al., 2015; Labati et al., 2011; Rota et al., 2015; Snousy et al., 2011). To the best of our knowledge, no study has attempted to propose integrated methodology for evaluating and benchmarking automated multiclass classification models for acute leukemia.

Two basic sets of criteria, namely, reliability and time complexity, are commonly used to evaluate and benchmark the multiclass classification modes of acute leukemia. Reliability contains a set of sub-criteria (true positive [TP], true negative [TN], false positive [FP], false negative [FN], ave-accuracy, precision μ, precision $_M$, recall $_M$, f-score and error rate) (Hossin & Sulaiman, 2015; Sokolova & Lapalme, 2009).

Snousy *et al.* considered that the main requirement that should be provided by the classification model to determine the best method is high accuracy (Snousy et al., 2011) Thus, nine classification models based on the accuracy criterion were compared in their study. Several studies (Campos et al., 2011; Jyoti Rawat et al., 2015; Salem et al., 2015; Wang et al., 2009; Yongqiang et al., 2015; Zhang & Xiaojuan, 2015) have adopted the classification accuracy criterion, despite the importance of other criteria, for evaluating and benchmarking the classification models (Cui et al., 2013; Krappe, Benz, et al., 2015; Mohapatra et al., 2016). However, the quality assessment of classification models for acute leukemia requires considerable attention. The same context (Snousy et al., 2011) states that other aspects must be considered in the evaluation processes. Rawat *et al*. mentioned that although accuracy is the most widely used metric, it considers each class of equal importance and neglects the differences among the types of classes (Jyoti Rawat et al., 2015); However, in real cases, especially in medicine, the distinction among certain classified classes is vital. Although various studies (Bhattacharjee & Saini, 2015; Chandra & Gupta, 2011; Singhal & Singh, 2014) have depended on TP, TN, FP, FN and sensitivity as key criteria for evaluation and benchmarking, other requirements that affect the classification performance have not been considered. Reference (Mishra et al., 2017) reported that the calculation complexity of dataset is a drawback, which is time consuming for classification. High computational cost slows down the classification (Rashid & Maruf, 2011). Misha *et al*. indicated that the dataset size should be considered in classification task because a large size will affect the processing time, namely, complexity time (Mishra et al., 2017).

The authors of (Ludwig et al. , 2015) stated that in the scope of cancer data analysis, speed and accuracy are the major aspects that must be considered in evaluating the efficiency of classification models. The main goal of successful classification tasks is to reduce the computational time while improving the classification accuracy (Saritha et al., 2016). Therefore, an integrated and comprehensive platform that covers all performance aspects in the evaluation and benchmarking of multiclass classification models of acute leukemia should be developed. This integrated methodology will serve as a tool to support the decisions of the administrations of medical organizations in evaluating and benchmarking available alternative models to determine the best one.

1.3 Research problem

The administrations of medical organizations have been facing difficulties in evaluating and benchmarking automated multiclass classification models for acute leukemia to determine the appropriate one, especially because no single model is superior to the rest (Agaian et al., 2014; Goutam & Sailaja, 2015) Furthermore, many models lack accuracy and computational efficiency (Mishra et al., 2017). Medical organizations have difficulty selecting the best model to use due to the diversity among available classification models.

In the information system perspective, benchmarking is a process of comparing the output of different systems for a given set of criteria to ensure the quality, improvement, contribution or performance of the new system (Trentesaux et

al., 2013). Benchmarking is conducted after the development of any classification model; this process aims to compare the new model with other similar models under the same conditions and metrics (Escalante et al., 2012). Most studies in this field have measured the multiclass classification model's performance by individually comparing its results with those of other previous models in accordance with a set of criteria (Al-Sahaf et al., 2013; Fan et al., 2008; Kim, 2009; Mohapatra et al., 2016; Saengsiri et al.. Figure 1.1 illustrates the problem statement configuration.

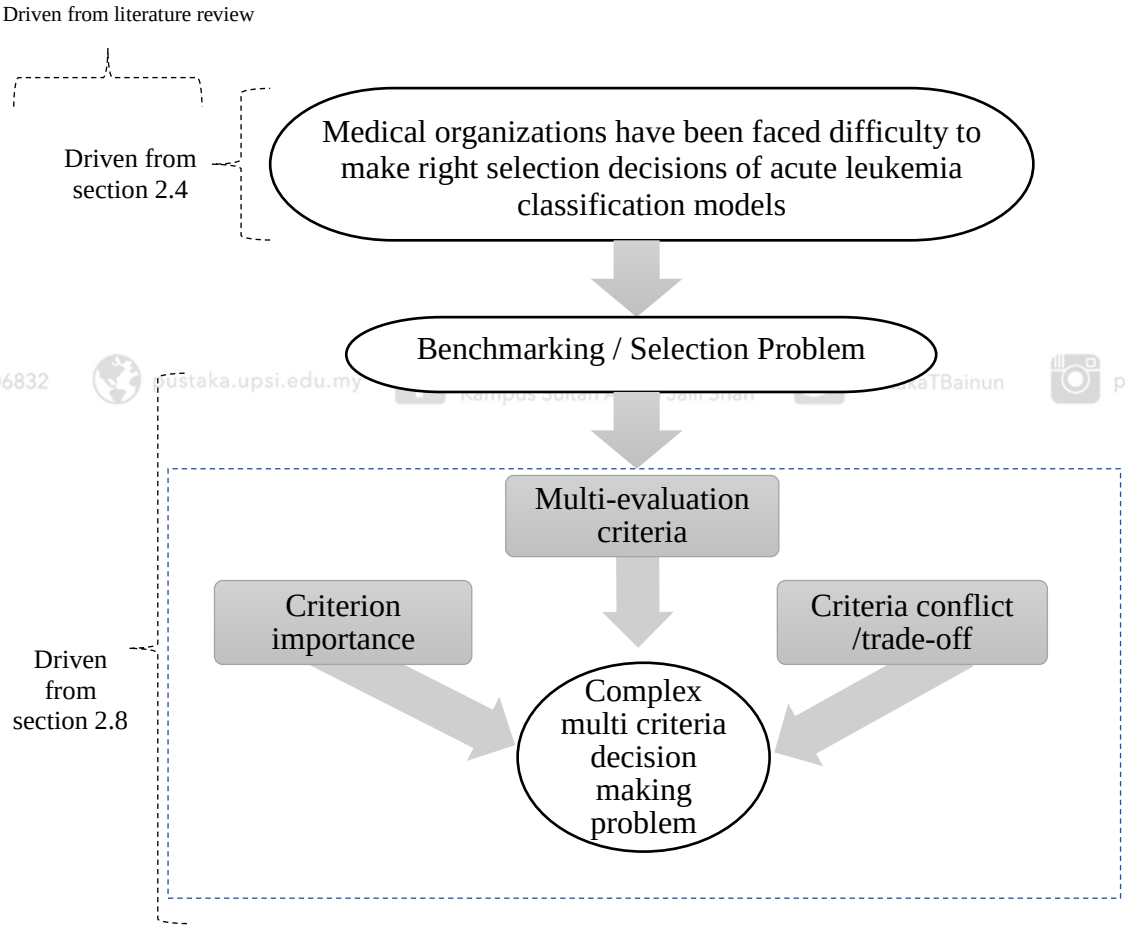

Figure 1.1 Problem Statement Configuration

The main requirements that must be considered in developing any multiclass classification models for acute leukemia are reliability and time complexity. Reliability should have a high rate, whereas time complexity for conducting the

output should be low (Saritha et al., 2016). However, these requirements are competing (Tai et al., 2011), i.e. high reliability cannot be simultaneously obtained with low time complexity. Thus, developers often focus on either increasing the reliability or decreasing the time complexity. In other words, if a highly reliable multiclass classification model is required, then we need to sacrifice on time, and vice versa.

The trade-off and conflict among the criteria are reflected on evaluation and benchmarking, thereby causing a conflict during comparison. Hence, the benchmarking process is affected, because benchmarking among multiple criteria is difficult with trade-off and conflict (Ganesh Kumar et al., 2012). To evaluate any classification model, the two main requirements, namely, reliability and time complexity, should be measured. However, the current comparison approach for the proposed and previous models in all reviewed studies disregards all evaluation and benchmarking criteria; it focuses on one aspect of the evaluation and ignores the rest, because it is not flexible enough to address the conflict or trade-off among various criteria (Snousy et al., 2011). Conflict and trade-off are considered the first issues in the evaluation and benchmarking of multiclass classification.

The second issue that affects evaluation and benchmarking is the importance of each criterion. The evaluation of multiclass classification models for acute leukemia involves a set of criteria, and the importance of each criterion varies in accordance with the objectives for which the model is developed. Hence, the

importance of one evaluation criterion may be increased in exchange for the low importance of another criterion (Goutam & Sailaja, 2015; He & Hui, 2009).

The third issue emerges when benchmarking among the classification models on the basis of multiple criteria and sub-criteria (Lu et al., 2012; Rosa, Magpantay et al., 2014; Yusen & Liangyun, 2010), this process is considered difficult due to the trade-off among the criteria and because each of them is important. Meanwhile, values of the reliability set of the criteria are criticised depending on the confusion matrix that contains four parameters, namely, TP, FP, TN and FN (Bhattacharjee & Saini, 2015; Kumar & Kumar Rath, 2015). The four parameters are prone to lose values in experiments, thereby affecting the values by all the other criteria. Despite the criticism with respect to these parameters, studies still use them for evaluating multiclass classification models (Dash, 2013; Lu et al., 2012; Rosa et al., 2014; Yusen & Liangyun, 2010).

The current evaluation and benchmarking tools are limited. These tools cannot entirely cover the requirements that should be measured by the multiclass classification model. They have limitations in calculating the overall parameters of the reliability group, comparing more than two classification methods and matching the classification methods. Furthermore, they cannot rank the models from best (with high performance) to worst (Rangra. & Bansal, 2014; Wahbeh et al., 2011; Yas et al., 2018). On this basis, the evaluation and benchmarking of multiclass classification models for acute leukemia are defined as a multi-criteria problems.

1.4 Research Questions

In order to set the direction of this research, the following research questions have been drawn up:

a) What is the available technology for evaluation of the automated classification tasks of acute leukemia?

b) What are the requirements needed for a benchmarking methodology for multiclass classification models of acute leukemia?

c) What are the criteria that have been used to evaluate and benchmark the multiclass classification models?

d) Is there any integrated methodology containing "evaluation criteria" and "multiclass classification models"?

e) What are the suitable techniques for develop benchmarking methodology for multiclass classification models of acute leukemia?

f) Are the results of the proposed benchmarking methodology valid?

1.5 Research Objectives

This study aims to develop benchmarking methodology for multiclass classification models using multi-criteria decision-making techniques. The objectives of this study are presented as follows:

1. To investigate the existing technology on evaluation for automated classification of acute leukemia and highlight the benchmarking tools' weaknesses.

2. To identify decision matrix based on multi-dimensional criteria for acute leukemia multiclass classification models.
3. To develop benchmarking methodology for acute leukemia multiclass classification models based on identified decision matrix.
4. To validate the proposed benchmarking methodology.

1.6 Relationship between Research Objectives, Research Questions and Research problem

Research questions are sketched to provide the direction and focusing of the research and the research objectives provide answers to the research questions. Table 1.1 presents the questions and their answered by objectives as well as it determines what part of research problem will be solved when each research objective achieved.

Table 1.1

Link Among Research Questions, Research Objectives and Research Problem

Research Questions	Research Objectives	Specific Problem	General problem
a) What is the available technology for evaluation the automated classification tasks of acute leukemia. b) What are the requirements needed to construct a benchmarking methodology for multiclass classification models of acute leukemia.	1. To investigate the existing technology on evaluation for automated classification of acute leukemia and highlight the benchmarking tools' weaknesses.	Multi Evaluation criteria problems.	Selection problem (Benchmarking)
c) What are the criteria that have been used to evaluation and benchmarking the multiclass classification models? d) Is there any integrated methodology containing "evaluation criteria" and "multiclass classification models"	2. To identify decision matrix based on multi-dimensional criteria for acute leukemia multiclass classification models.		

(*Continue*)

Table 1.1(*Continued)*

Research Questions	Research Objectives	Specific Problem	General problem
e) What are the suitable techniques for develop benchmarking methodology for multiclass classification models of acute leukemia.	3.To develop benchmarking methodology for multiclass classification models based on based on identified decision matrix	-Trade off criteria and Conflicting criteria. -Importance of criteria. - Multi evaluation criteria	
f) Are the results of proposed benchmarking methodology valid?	4. To validate the proposed benchmarking methodology. .		

1.7 Scope of the study

The scope of this research is defined by the following considerations:

1. This research focuses on the development benchmarking methodology for multiclass classification models based on multi-criteria decision analysis of acute leukemia.
2. Development multiclass classification models of acute leukemia is not the main issue of this study, they will be for proof of concept to our proposed methodology.

Figure 1.2. illustrate the general view for our research and view representing the research method, research type, and research domain.

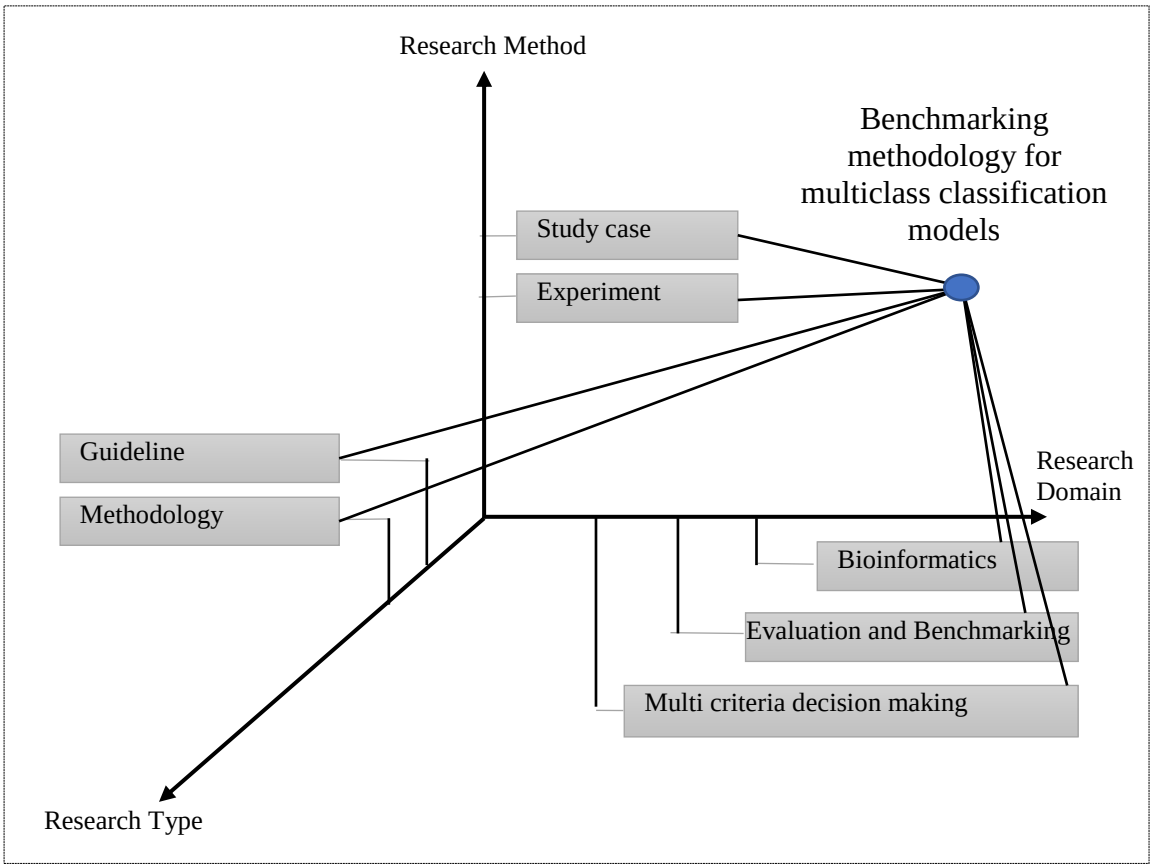

Figure 1.2 Research Scope

This study is a multi-disciplinary involving benchmarking methodology for multiclass classification models in order to evaluate the multiclass classification models that considered one of the biomedical models. The study is designed to address the benchmarking/selection problem of multiclass classification models. In the case study, multiclass classification models of acute leukemia are used in experiments to generate the data that used to proof of concept of our proposed methodology.

The outcomes of the research indicate the research type. Two outputs are from this study. One is a methodology performed via several steps that improve the process

of evaluation and benchmarking of multiclass classification models. The other is a complete guideline to evaluate and benchmarking multiclass classification models.

The integrated MCDM method used to test (evaluation and benchmarking) the performance of multiclass classification in order to improve the decision of medical organization. Therefore, our research belongs to the information system domain.

1.8 Motivation

Accurate and rapid medical diagnosis is essential in order to provide the most effective treatment option, especially in serious diseases such as acute leukemia (Singhal & Singh, 2014). An accurate classification of cancer has great value in providing better treatment and fast response, so we need a reliable, precise automated solutions for classification of acute leukemia (Agaian et al., 2014; Mohapatra et al., 2010; Srisukkham et al., 2017). There are multiple and different automated solutions for classification of acute leukemia; however, it is difficult to take a right decision to determine which one is more reliable and accurate in classification. So, the main motivation is to enable medical organizations to make sound decisions concerning the selection the proper classification models. Choosing the right automated classification models that give high performance and accurate results would increase satisfaction and efficiency, improved medical care operations and quality of care(Bagasjvara et al., 2016; Escalante et al., 2012; Singh et al., 2016). This study will be beneficial to the hospitals and cancer treatment centers in its decision to select the best automated

models for acute leukemia classification. This study can incentive for researchers and scholars to propose automated classification solutions that support medical organizations in decision making, as well as motivate them to further explore the measurement and evaluation area in the medical industry. There exist only a few publications on the topic of empowering of health organizations to make a sound decision regarding the selection of automated multiclass classification model for acute leukemia, and what is related of their evaluation and benchmarking using MCDM.

1.9 Significance of the study

1.9.1. Practical significance of the study

Practically, through the proposed benchmarking methodology for multiclass classification models of acute leukemia, hospitals and cancer treatment centers will be able to choose the most reliable automated models to classify the acute leukemia. Thus, they will be able to provide a precise and reliable for acute leukemia classification and provide the treatment services, thus that's enhances the performance of health organizations and achieve patient confidence(Bagasjvara et al., 2016; Goutam & Sailaja, 2015). Through the proposed benchmarking methodology, the decisions of health organizations regarding the choice of a multiclass classification models will be accurate and based on scientific method that will be developed and tested according to the sound scientific basis.

1.9.2. Theoretical importance the study

This study contributes through adopting the systematic literature review approach to provide an overview of existing information and proofs with respect to automated classification tasks and their evaluation and benchmarking approach, as well as to highlight the trends of research work on this topic. This study also contributes to filling the lack of research in this research area. The proposed taxonomy of the related literature in this study can bring several benefits as well, imposes a sort of organization on the mass of publications, sort out those different works into a meaningful, manageable and coherent layout, and provides all researchers with important insights into the subject field in several ways, others importance of proposed taxonomy, it outlines the potential directions of research in the field, it can reveal gaps in researches, and mapping the literature on automated methods of acute leukemia detection and classification into distinct categories highlights weak and strong features in terms of research coverage (Hussain et al., 2015). In addition, this study provides a guide to the most important criteria should to be adopted to evaluate multi-class classification models.

1.10 Main Terminology

This section presents description of the main terms of this study

- Multiclass classification means a classification task based on machine learning techniques with more than two classes, in our context, classify the gens data sample into three leukemia categories, namely, AML, ALL-T, ALL-B.

- Benchmarking process means comparing the output of different systems for a given set of criteria to ensure the quality, improvement, contribution, or performance of the new system.
- Multi criteria decision analysis is an umbrella term to describe a collection of formal approaches, which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter.

1.11 Organization of research

This study is composed of five chapters, figure 1.3 illustrate the structure of study. The background of research, research problem, research questions, research objective, relationship between research questions, research objective with research problem, research scope and motivation with significant of research is provided in chapter one. Thus, the remaining of this research are organized as follows:

Chapter two: "Literature review". In this chapter, the existing literature on evaluation and benchmarking approaches for acute leukemia multiclass classification models is discussed. Through this chapter, the main criteria for evaluation and benchmarking are identified and described in details. Also, existing methods of evaluation and measurement is discussed with related problems and issues that they are suffered from. This chapter also include the theoretical background of multi criteria decision making (MCDM) in details, presents the popular MCDM methods, and explain the main two MCDM methods: Best-Worst Method and VIKOR method. Furthermore, it

explains the useful techniques that enable them to take decision for the multi-criteria problems. The main purpose of this chapter is figure out the research gap and challenges as well as to propose the recommended solution.

Chapter three: "Research Methodology". This chapter describes the requirements for development the proposed benchmarking methodology for multiclass classification models as well as the phases followed. The methodology is designed in four key phases, namely, investigation phase, identification phase, development phase and validation phase. Through the phases, this chapter will present in detail how the four research objectives will be achieved.

Chapter four "Results and Discussion" This chapter presents the results and discussion of the benchmarking methodology for multiclass classification models carried out. The chapter demonstrates how the results of the proposed methodology resolve the problems mentioned in the problem statements, also this chapter presents the results of validation process.

Finally, **chapter five** "conclusion and future work" concludes and summarizes the research contributions made. Moreover, research limitation, further research proposals and the conclusion are reported.

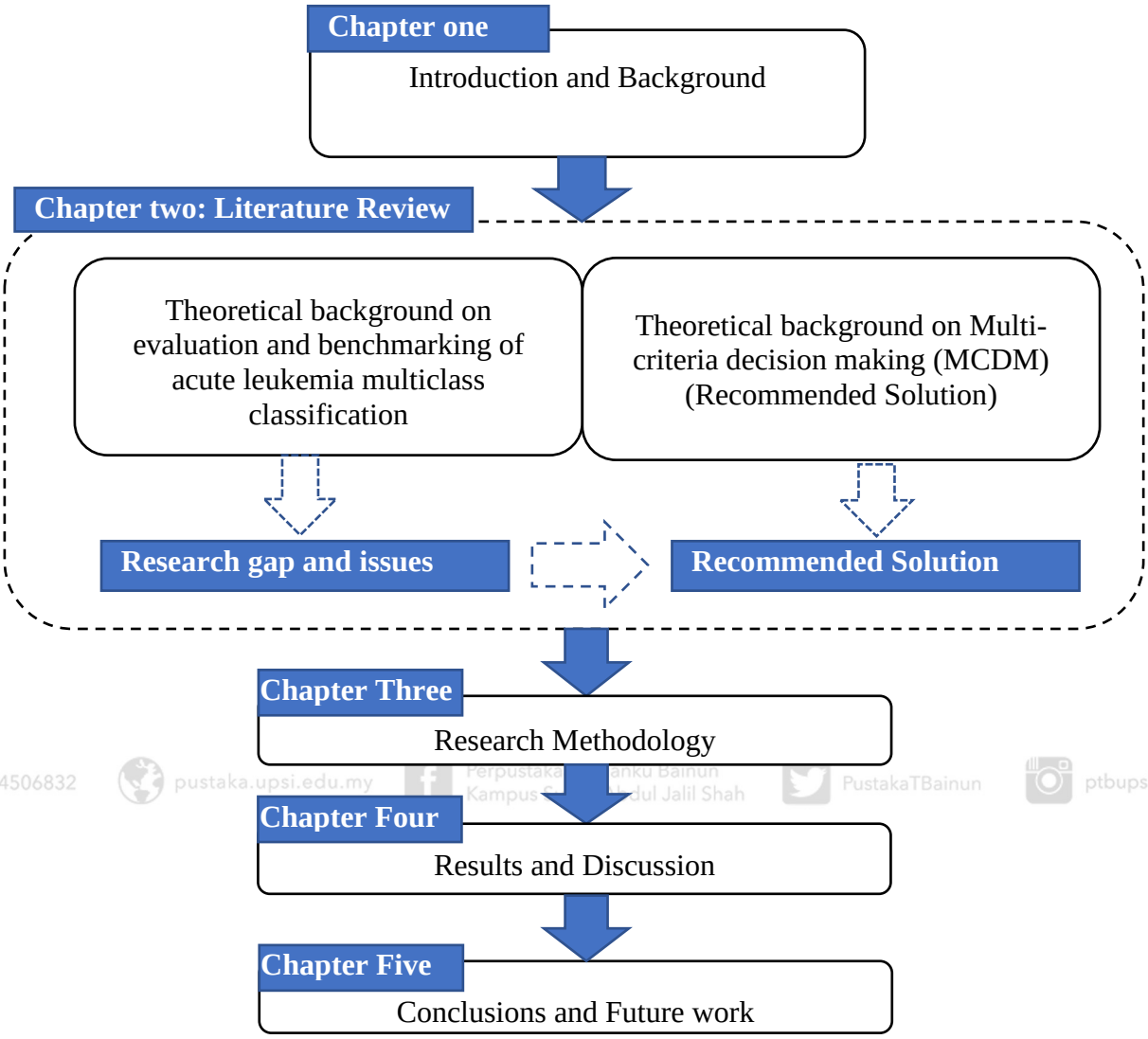

Figure 1.3 Structure of the Dissertation