

**BORANG PENGESAHAN STATUS TESIS**

JUDUL: **FETAL WEIGHT PREDICTION USING NEURAL ENSEMBLE  
BASED C4.5 FOR LOW BIRTH WEIGHT FETUSES**

SESI PENGAJIAN: **2 - 2009/2010**

Saya **LOO KIEN LIM**

(HURUF BESAR)

mengaku membenarkan tesis (PSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. \*\* Sila tandakan (/)

SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

TERHAD (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

TIDAK TERHAD

  
\_\_\_\_\_  
(TANDATANGAN PENULIS)  
Alamat tetap: **BLOK A16, LOT B6,  
KONDOMINIUM LPPB,  
BATU 3.5, JLN. LABUK,  
90000 SANDAKAN,  
SABAH, MALAYSIA.**

Tarikh: 30 June 2010

  
\_\_\_\_\_  
(TANDATANGAN PENYELIA)  
**DR. CHOO YUN HUOY**  
Nama Penyelia

Tarikh: 30/6/2010

CATATAN: \* Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM)

\*\* Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.

**FETAL WEIGHT PREDICTION USING NEURAL ENSEMBLE BASED C4.5  
FOR LOW BIRTH WEIGHT FETUSES**

**LOO KIEN LIM**

**This report is submitted in partial fulfillment of the requirements for the  
Bachelor of Computer Science (Artificial Intelligence)**

**FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY  
UNIVERSITI TEKNIKAL MALAYSIA MELAKA  
2010**

**DECLARATION**

I hereby declare that this project report entitled

**FETAL WEIGHT PREDICTION USING NEURAL ENSEMBLE BASED C4.5  
FOR LOW BIRTH WEIGHT FETUSES**

is written by me and is my own effort and that no part has been plagiarized

without citations.

STUDENT

:   
\_\_\_\_\_  
(LOO KIEN LIM) Date: 30 June 2010

SUPERVISOR

:   
\_\_\_\_\_  
(DR. CHOO YUN HUOY) Date: 30/6/2010

## **DEDICATION**

To my beloved parents, Mr. Loo Geck Meng and Mrs. Siaw Fung Chiew, for their  
seems less expression of love and fully support...

To my supervisor, Dr. Choo Yun Huoy, for making it all worthwhile...

## ACKNOWLEDGEMENTS

First of all, I would like to express my sincere appreciation to my respectable supervisor, Dr. Choo Yun Huoy for her valuable advice and assistance in choosing the topic, recommendations for solutions and materials review.

Besides that, I would like to thank Universiti Teknikal Malaysia Melaka for offering PSM (*Projek Sarjana Muda*) as a compulsory subject for UTeM students in order to fulfill the degree. It is also a note of thanks especially dedicated for PSM committee members on their kindness in organizing briefing, seminar and talk to PSM students.

Also, I would like to convey my sincere thanks to my evaluator, Dr. Burhannudin bin Mohd Aboobaider for his reasonable evaluation and also valuable advice.

Moreover, I also would like to express my sincere thanks to my faithful course mates for their willingness to share their ideas, knowledge and resources.

Last but not least, I would also like to thank my beloved parents who have been giving me support and motivation throughout my project.

## ABSTRACT

With the continually increasing of infants' mortality rates and birth defect rates, Low Birth Weight (LBW) is of concern for everyone. A woman's behavior during pregnancy such as smoking habits can greatly alter the chance of carrying infant in abnormal term. LBW fetus is baby with a weight of less than 2,500 grams at birth. Characteristics of baby with LBW include born prematurely and small body for their age. Because of the problems with the placenta and the mother's health particularly during pregnancy, the infant has less time to grow and gain weight in the mother's uterus. It can also caused by environmental factors and inheritance genetics. Consequences, LBW infants face a higher risk of death within the first year of life and have higher rates of disability and disease than other normal infants. Hence, the purpose of this project is to purpose a classification technique, Neural ensemble Based C4.5 (NeC4.5) on prediction of LBW fetuses through developing a classification application while compare the result analysis with Decision Tree C4.5 and RBF Neural Network. This thesis describes the analysis, design of NeC4.5 and result analysis. Basically, this application involves three modules which is data loading, data classification and result analysis. Object Oriented Analysis and Design has been chosen as a methodology for this project and will be implemented along the application development process. After carried 10x10 cross-validation of classification for NeC4.5, result analysis in term of mean Accuracy and mean F-measure shown that NeC4.5 has a best overall performance than the other two techniques due to it is a hybrid technique where its generalization ability can be better than Decision Tree C4.5 and its good comprehensibility. Hence, NeC4.5 is a good classifier towards solving LBW problem. However, the execution time performance still needs to improve in the future.

## ABSTRAK

Dengan peningkatan mendadak bagi kadar kematian dan kecacatan kelahiran bayi, berat lahir rendah adalah dititikberat oleh masyarakat. Tabiat bagi wanita hamil secara tidak langsung adalah disyaki yang mengubah peluang pembawaan bayi yang luar biasa. Akibatnya, kejadian berat lahir rendah bayi dengan keberatan kurang daripada 2,500 grams selepas lahir. Biasanya, ciri-ciri bayi yang lahir berat rendah adalah sama ada lahir sebelum cukup waktunya atau memiliki badan size yang kecil bagi umur mereka disebabkan masalah plasenta, kesihatan ibu terutamanya semasa tempoh kehamilan, kekurangan masa perkembangan dan kedapatan berat secukupi bagi bayi dalam ibu rahim, faktor-faktor persekitaran, genetik warisan dan sebagainya. Justeru, hal kejadian ini menyebabkan bayi menghadapi risiko yang tinggi meninggal dunia dalam tempoh satu tahun selepas kelahiran dan kekurangupaya serta menghadapi penyakit berbanding bayi yang biasa. Dengan itu, tujuan projek ini adalah mencadangkan satu teknik pengkelasan yang dinamakan Neural ensemble Based C4.5 (NeC4.5) bagi ramalan berat lahir rendah bayi melalui perbanguan satu aplikasi pengkelasan sambil membandingkan keputusan analisisnya dengan dua teknik pengkelasan iaitu Decision Tree C4.5 dan RBF Neural Network. Tesis ini menerangkan analisa, rekacipta bagi NeC4.5 dan keputusan analisa. Biasanya, aplikasi ini terdapat tiga modul iaitu muatan data, pengkelasan data dan keputusan analisa. Teknik Analisa dan Rekacipta Berorientasikan Objek dipilih sebagai metodologi bagi projek ini dan akan dilaksanakan dalam sepanjang proses pembangunan aplikasi ini bagi memastikan matlamat projek ini tercapai. Selepas menjalani *10x10 cross-validation* bagi pengkelasan NeC4.5, keputusan analisa dalam bentuk julat ketepatan (*mean Accuracy*) dan julat pengukuran F (*mean F-measure*) menunjukkan bahawa NeC4.5 mempunyai keseluruhan prestasi yang terbaik berbanding dengan dua teknik yang lain itu disebabkan ia merupakan teknik kacukan di mana kebolehan pembiakannya lebih baik daripada Decision Tree C4.5 dan pemahamannya yang baik. Maka, NeC4.5 adalah pengkelas yang baik bagi menyelesaikan masalah LBW. Akan tetapi, prestasi masa perlaksanaan masih perlu diperbaiki pada masa depan.

## TABLE OF CONTENTS

CHAPTER	SUBJECT	PAGE
	<b>PROJECT TITLE</b>	<b>i</b>
	<b>DECLARATION</b>	<b>ii</b>
	<b>DEDICATION</b>	<b>iii</b>
	<b>ACKNOWLEDGEMENTS</b>	<b>iv</b>
	<b>ABSTRACT</b>	<b>v</b>
	<b>ABSTRAK</b>	<b>vi</b>
	<b>TABLE OF CONTENTS</b>	<b>vii</b>
	<b>LIST OF TABLES</b>	<b>xi</b>
	<b>LIST OF FIGURES</b>	<b>xii</b>
	<b>LIST OF ABBREVIATIONS</b>	<b>xiii</b>
	<b>LIST OF ATTACHMENTS</b>	<b>xiv</b>
<b>CHAPTER I</b>	<b>INTRODUCTION</b>	<b>1</b>
	1.1 Project Background	1
	1.2 Problem Statements	2
	1.3 Objectives	3
	1.4 Scope	3
	1.5 Project Significance	4
	1.6 Expected Output	4
	1.7 Conclusion	4
<b>CHAPTER II</b>	<b>LITERATURE REVIEW AND PROJECT METHODOLOGY</b>	<b>6</b>
	2.1 Introduction	6
	2.2 Facts and Findings	7
	2.2.1 The Review of Classification in	8

	Medical Research	
	2.2.2 Technique	17
2.3	Project Methodology	18
2.4	Project Requirements	23
	2.4.1 Software Requirement	23
	2.4.2 Hardware Requirement	24
	2.4.3 Other Requirements	24
2.5	Project Schedule and Milestones	25
2.6	Conclusion	25
<b>CHAPTER III</b>	<b>ANALYSIS</b>	26
3.1	Introduction	26
3.2	Problem analysis	26
	3.2.1 WEKA Data Mining Toolkit	27
	3.2.2 CART software	28
3.3	Requirement analysis	29
	3.3.1 Functional Requirement	29
	3.3.2 Non-functional Requirement	32
	3.3.3 Others Requirement	34
3.4	Conclusion	35
<b>CHAPTER IV</b>	<b>DESIGN AND IMPLEMENTATION</b>	36
4.1	Introduction	36
4.2	High-Level Design	37
	4.2.1 Raw Data/Input	37
	4.2.2 System Architecture	38
	4.2.2.1 High-level Class Diagram	38
	4.2.3 User Interface Design	38
	4.2.3.1 Navigation Design	39
	4.2.3.2 Input Design	39
	4.2.3.3 Technical Design	39

	4.2.3.4 Output Design	40
	4.2.3.5 Prototype User Interface	40
	4.2.4 Non-database Design	41
	4.2.4.1 Conceptual and Logical Design	41
4.3	Detailed Design and Development Setup	41
	4.3.1 Software or Hardware Design and Development Setup	42
	4.3.2 Physical Design	44
4.4	Conclusion	44
<b>CHAPTER V</b>	<b>TESTING</b>	<b>45</b>
5.1	Introduction	45
5.2	Test Plan	46
	5.2.1 Test Organization	46
	5.2.2 Test Schedule	46
5.3	Test Strategy	47
	5.3.1 Classes of Tests	47
	5.3.1.1 Cross-Validation	48
	5.3.1.2 Paired T-test	48
	5.3.1.3 Validation Testing	49
5.4	Test Implementation	49
	5.4.1 Test Description	49
	5.4.1.1 Data File Type Testing	50
	5.4.1.2 User Input Testing	50
	5.4.1.3 Paired T-test	50
	5.4.2 Test Data	51
5.5	Conclusion	51
<b>CHAPTER VI</b>	<b>RESULT ANALYSIS</b>	<b>52</b>
6.1	Introduction	52

6.2	Result Analysis	52
6.3	Comparison among Classifier Models and Ranking	56
6.4	Conclusion	57
<b>CHAPTER VII</b>	<b>PROJECT CONCLUSION</b>	<b>58</b>
7.1	Observation on Weaknesses and Strengths	58
7.2	Proposition for Improvement	59
7.3	Contribution	60
7.4	Conclusion	61
	<b>REFERENCES</b>	<b>62</b>
	<b>BIBLIOGRAPHY</b>	<b>64</b>
	<b>ATTACHMENTS</b>	<b>65</b>

## LIST OF TABLES

TABLE	TITLE	PAGE
2.1	Proposed Survivability Class Instances	12
2.2	Performance of classification techniques with different measurements	16
3.1	Software Requirements for LBW Classification Application	34
3.2	Hardware Requirements for LBW Classification Application	35
4.1	Data Loading Input Design	39
4.2	Data Loading Output Design	40
4.3	Data Classification Output Design	40
5.1	Test Schedule	46
6.1	Total Average Accuracy of Whole Folds for Three Classifiers	53
6.2	Total Average F-measure of Whole Folds for Three Classifiers	55
6.3	Ranking of best classifier based on Average Accuracy	57
6.4	Ranking of best classifier based on Average F-measure	57

## LIST OF FIGURES

DIAGRAM	TITLE	PAGE
2.1	Pre-classification on breast dataset	11
2.2	Information Gain measure	12
2.3	Bagging approach	14
2.4	C4.5 Rule-PANE	14
2.5	Bagging approach	17
2.6	NeC4.5 algorithm	18
3.1	Algorithm for NeC4.5 classification application	30
3.2	Algorithm for NeC4.5 workflow	31
3.3	Algorithm for result analysis and comparison	32
3.4	Accuracy calculation	32
3.5	Confusion Matrix with Description	33
3.6	Calculation of Precision, Recall and F-measure	34
6.1	Graph Representation Average Accuracy of Whole Folds for Three Classifiers	54
6.2	Graph Representation Average F-measure of Whole Folds for Three Classifiers	56

## LIST OF ABBREVIATIONS

ABBREVIATION	-	NAME
ANE/ANNE	-	Artificial Neural Network Ensemble
ANN	-	Artificial Neural Network
BPNN	-	Back-propagated Neural Network
C4.5	-	Decision Tree C4.5
C4.5 Rule-PANE	-	C4.5 Rule preceded by Artificial Neural Network Ensemble
CART	-	Classification and Regression Tree software
COD	-	Cause of Death
FWE-LBWF	-	Fetal Weight Estimation for Low Birth Weight Fetuses
GUI	-	Graphical User Interface
LBW	-	Low Birth Weight
LR	-	Logistic Regression
NeC4.5	-	Neural Ensemble Based C4.5
OOA	-	Object-oriented Analysis
OOAD	-	Object-Oriented Analysis and Design
OOD	-	Object-oriented Design
RAD	-	Rapid Application Development
RBFN	-	RBF Neural Network
ROC	-	Receiving Operating Curves
SIDS	-	Sudden Infant Death Syndrome
STR	-	Survival Time Record
SVM	-	Support Vector Machine
UML	-	Unified Modeling Language
VSR	-	Vital Status Recode
WEKA	-	Waikato Environment for Knowledge Analysis

## LIST OF ATTACHMENTS

ATTACHMENT	TITLE	PAGE
<b>A</b>	<b>Project Schedule and Milestones</b>	
	<b>A.1 Project Activities</b>	<b>66</b>
	<b>A.2 Gantt Chart</b>	<b>69</b>
<b>B</b>	<b>Current System Activity Diagram</b>	
	<b>B.1 Activity Diagram for WEKA</b>	<b>75</b>
	<b>B.2 Activity Diagram for CART</b>	<b>76</b>
<b>C</b>	<b>NeC4.5 Classification Application Activity Diagram</b>	
	<b>C.1 Activity Diagram for NeC4.5 classification application</b>	<b>78</b>
	<b>C.2 Activity Diagram for NeC4.5 workflow</b>	<b>79</b>
	<b>C.3 Activity Diagram for result analysis and comparison</b>	<b>80</b>
<b>D</b>	<b>Dataset</b>	
	<b>D.1 Raw Data</b>	<b>82</b>
	<b>D.2 Preprocessed Data</b>	<b>82</b>
<b>E</b>	<b>System Architecture</b>	
	<b>E.1 Classification of LBW fetuses' overview</b>	<b>84</b>
	<b>E.2 NeC4.5 classification application tier approach</b>	<b>85</b>
	<b>E.3 Classification of LBW fetuses with NeC4.5</b>	<b>86</b>
<b>F</b>	<b>High level Class Diagram</b>	<b>88</b>
<b>G</b>	<b>NeC4.5 classification Navigation Design</b>	<b>90</b>
<b>H</b>	<b>User Interface Design</b>	
	<b>H.1 User interface design for NeC4.5 Classification (Welcome interface)</b>	<b>92</b>

	<b>H.2 User interface design for NeC4.5 classification</b>	<b>93</b>
	<b>H.3 User interface design for Data Loading (Data File Chooser)</b>	<b>93</b>
	<b>H.4 User interface design for Data Visualization</b>	<b>94</b>
	<b>H.5 User interface design for message box (if not a .arff or .csv file type)</b>	<b>94</b>
	<b>H.6 User interface design for message box (if wrongly setup of fold and seed numbers)</b>	<b>95</b>
	<b>H.7 User interface design for message box (if default setup of fold and seed numbers)</b>	<b>95</b>
	<b>H.8 User interface design for Results Classification (Fold-by-Fold)</b>	<b>96</b>
	<b>H.9 User interface design for Results Classification (Average of whole folds)</b>	<b>97</b>
	<b>H.10 User interface design for Measurements Information/Descriptions</b>	<b>98</b>
<b>I</b>	<b>Logical and Conceptual Design of NeC4.5 Classification application</b>	<b>100</b>
<b>J</b>	<b>Algorithm for NeC4.5 Classification Application Function</b>	
	<b>J.1 Algorithm for Data Loading Function</b>	<b>102</b>
	<b>J.2 Algorithm for DataVisualization Function</b>	<b>107</b>
	<b>J.3 Algorithm for Data Classification Function</b>	<b>108</b>
<b>K</b>	<b>Physical Design of NeC4.5 Classification application</b>	<b>112</b>
<b>L</b>	<b>Testing</b>	
	<b>L.1 Data File Type Testing</b>	<b>114</b>
	<b>L.2 User Input Testing</b>	<b>115</b>
	<b>L.3 Paired T-test</b>	<b>116</b>
	<b>L.4 Test Data for Data File Type Testing</b>	<b>116</b>
	<b>L.5 Test Data for User Input Testing</b>	<b>117</b>
	<b>L.6 Paired T-test for Low Birth Weight data</b>	<b>118</b>
	<b>L.7 Paired T-test for Breast Cancer data</b>	<b>119</b>
<b>M</b>	<b>Average Accuracy and Average F-measure for whole Random Seeds of 1-10 folds Cross-validation</b>	
	<b>M.1 Average Accuracy of Fold 1</b>	<b>121</b>

<b>M.2 Average Accuracy of Fold 2</b>	<b>121</b>
<b>M.3 Average Accuracy of Fold 3</b>	<b>122</b>
<b>M.4 Average Accuracy of Fold 4</b>	<b>122</b>
<b>M.5 Average Accuracy of Fold 5</b>	<b>123</b>
<b>M.6 Average Accuracy of Fold 6</b>	<b>123</b>
<b>M.7 Average Accuracy of Fold 7</b>	<b>124</b>
<b>M.8 Average Accuracy of Fold 8</b>	<b>124</b>
<b>M.9 Average Accuracy of Fold 9</b>	<b>125</b>
<b>M.10 Average Accuracy of Fold 10</b>	<b>125</b>
<b>M.11 Average F-measure of Fold 1</b>	<b>126</b>
<b>M.12 Average F-measure of Fold 2</b>	<b>126</b>
<b>M.13 Average F-measure of Fold 3</b>	<b>127</b>
<b>M.14 Average F-measure of Fold 4</b>	<b>127</b>
<b>M.15 Average F-measure of Fold 5</b>	<b>128</b>
<b>M.16 Average F-measure of Fold 6</b>	<b>128</b>
<b>M.17 Average F-measure of Fold 7</b>	<b>129</b>
<b>M.18 Average F-measure of Fold 8</b>	<b>129</b>
<b>M.19 Average F-measure of Fold 9</b>	<b>130</b>
<b>M.20 Average F-measure of Fold 10</b>	<b>130</b>
<b>N User Manual</b>	<b>132</b>

## **CHAPTER I**

### **INTRODUCTION**

#### **1.1 Project Background**

With the continually increasing of infant mortality rates and birth defect rates, Low Birth Weight (LBW) is of concern for everyone. A woman's behavior during pregnancy including diet, smoking habits, and received prenatal care can greatly alter the chances of carrying the infant to term, consequently, of delivering an infant of abnormal birth weight called as LBW.

A baby with a weight of less than five pounds (2,500 grams) at birth is classified as a LBW baby. Characteristics of baby with LBW include born prematurely and small body for their age. Due to the problems with the placenta and the mother's health particularly during pregnancy, the infant has less time to grow and gain weight in the mother's uterus. Furthermore, it can also be caused by environmental factors, inheritance genetics or other factors. Consequently, LBW infants face a higher risk of death within the first year of life and have higher rates of disability and suffer several diseases such as Sudden Infant Death Syndrome (SIDS) than other normal infants. Hence, we must concern on this issue and prevent it from continuously occurring by predicting this occurrence of LBW infants.

Moreover, it is difficult to find the relationship between infants and maternal in the way to solve this problem. This relationship is important for pregnant in the way to determine whether a maternal carry a LBW fetuses. If that relationship is

strong, then the maternal health and her lifestyle is more concern in the way for successfully delivery of normal fetus. So, in order to prevent delivery of LBW fetus, this relationship must be known.

Furthermore, the relationship between LBW fetuses and lifestyle of normal infants also concern to be important. This relationship is vital as it determine how closer of LBW fetuses with normal infants lifestyle. If that relationship is strong, thus LBW case must take a lot of concern as it can affect the lifestyle of infants in growing process. More seriously, LBW can decrease population rates.

Currently, researcher has been conducting the classification of LBW data by using Logistic Regression (LR), tree classification, Random Forest, and etc. However, all of these techniques are statistical techniques, the role of soft computing techniques still haven't realized.

## **1.2 Problem Statements**

It is hard to predict the causes of this problem (LBW). The delivery of LBW fetuses probably caused from various factors whether there is factor environments, mother's health or factor genetic. Hence, there is a conflicting task to deal with this problem without knowing the causes of LBW. Consequence, the infant mortality rates and birth defect rates are continually increasing over the year.

Besides that, there is also no researcher use this hybrid soft computing technique, NeC4.5 on the prediction of LBW fetuses. According to Iacus and Porro (2006), most of them use statistical techniques on this prediction. Hence, the efficiency of soft computing technique towards this prediction has not been realised.

Moreover, medical doctor is still finding the factors of LBW fetuses. They want to know the exact causes of LBW fetuses in order to decrease infant mortality rates and birth defect rates.

### 1.3 Objectives

The main objectives of the proposed project are as follows:

- To propose classification technique on prediction of LBW fetuses.
- To compare proposed classification techniques with two others benchmark classification techniques in the prediction of LBW fetuses.
- To develop classification application based on the proposed classification technique.

### 1.4 Scope

The scope for this project is going to propose a classification technique which allows users to perform classification task on LBW dataset collected from Baystate Medical Center, Springfield, Massachusetts. Besides that, I also performed classification over LBW dataset by using proposed hybrid soft computing technique, Neural ensemble Based C4.5 (NeC4.5). Then, compare with two others benchmark classification techniques which are Decision Tree C4.5 (C4.5) and RBF Neural Network (RBFN). Moreover, it is also classify data with appropriateness setting of test options. For this classification of LBW dataset, it uses 10-fold cross validation to train the data. It is also compare the classifiers by using suitable evaluation performance measurements in order to determine which classifier has a best overall performance towards LBW fetuses meanwhile time performance are not in term of measurement. In this project, the average accuracy and the average F-measure of classifiers are assessed. Then, a Paired T-test is used to compare the performance of proposed classification technique with two others benchmark classification techniques on the prediction of LBW fetuses. Furthermore, this project going to use Java Programming language to code the proposed classification technique (NeC4.5). Lastly, the target user for this project is medical doctor (Assistance) who works in Child Department of hospital, researcher who related with this field and data miner.

## 1.5 Project Significance

The medical doctor can perceive whether proposed classification technique available on the prediction of LBW fetuses. Moreover, the prediction of fetal weight towards LBW fetuses may more easily by using this hybrid soft computing classification technique if compared with common existing classification techniques. The medical doctor also can find the relationship between infant and maternal whether it is strong, weak or no relation. If there is a strong relationship found, thus an advice or suggestion must be given to aware pregnant woman. Furthermore, researcher able analyzes the result after classification. Hence, this task can assist pregnant mother to predict whether her fetus(s) is/are belong to LBW category.

## 1.6 Expected Output

A result analysis of the proposed classification technique performance towards the prediction of LBW fetuses. Besides that, there is also a comparison analysis of proposed classification technique with two others benchmark classification techniques in the prediction of LBW fetuses. Moreover, the best classification technique on LBW fetuses' prediction of which either soft computing or statistical classification techniques is revealed through rank ordering.

## 1.7 Conclusion

As a conclusion, this project provides an opportunity to propose a classification technique on the prediction of LBW fetuses. It also reveals that whether proposed classification technique provides best performance towards LBW fetuses' prediction and a comparison analysis of proposed classification technique with two others benchmark classification techniques in the prediction of LBW fetuses. Besides that, a best classification technique either soft computing or

statistical classification techniques on LBW fetuses prediction is revealed. After that, the literature review which related to the LBW classification and methodology for this application are described in Chapter II.

## CHAPTER II

### LITERATURE REVIEW AND PROJECT METHODOLOGY

#### 2.1 Introduction

Literature review is an essential part of many projects as it gives us a valuable opportunity to read more on the subject that relevant to the project and observes how others have approached on the proposed area. Hence, the literature review on this chapter is conducted to gather more information to be applied in the way of purposing classification technique on prediction of LBW fetuses while compare among others benchmark classification techniques in the prediction of LBW fetuses. The discussion surrounds the classifiers for prediction task that related to this project such as Neural ensemble Based C4.5 (NeC4.5) by Zhi-Hua Zhou and Yuan Jiang (2004), Decision Tree C4.5 (C4.5) by Ross Quinlan (1993) and RBF Neural Network (RBFN) by He and Lapedes (1991). Then, the evaluation performance measurements of those classifiers in the past research mostly used Receiving Operating Curves (ROC), sensitivity, specificity analyses and other performance measurements to assess the performance of those classifiers. All of the classifiers stated above have its own method or function for prediction: NeC4.5 as a hybrid soft computing classification technique with integrated of Neural Network ensemble into Decision Tree C4.5 which perform prediction by first trains a Neural Network ensemble. Then, generate new training set from the trained ensemble. After that, the new training set is trained using Decision Tree C4.5. Meanwhile, Decision Tree C4.5 which in a tree structure form perform classification using a tree hierarchy moving from the tree's root to leaf node whereas RBFN uses a radial neurons function for prediction. There is no any previous work has been done on

implementation of NeC4.5 technique to handle with LBW fetuses prediction. Thus, this project aims going to propose NeC4.5 technique on classification of LBW dataset in other word is LBW fetuses' prediction.

Besides that, the methodology that is going used for this project is Object-Oriented Analysis and Design (OOAD) Technique. Generally, OOAD is a software engineering approach which takes a group of objects interaction or known as relationship to model a system. Generally, this methodology is divided into two part which are Object-oriented Analysis (OOA) analyzes the functional requirements of a system by applying object-modeling techniques and Object-oriented Design (OOD) produces specifications of implementation by elaborating the analysis models. However, in order to implement this project successfully, it is require undergoes Data Preprocessing for the purpose of data cleaning, missing or inconsistent data handling in order to ensure data integrity and validity. After that, the preprocessed data will be classified into two classes for the purpose of prediction LBW fetuses. Then, the classification result will be analyzed through assessment of mean Accuracy and mean F-measure. Lastly, comparison among classifiers through Paired T-test to reveal which classifier has performance towards solving LBW problems.

## **2.2 Facts and Findings**

Various materials have been retrieved from internet, data source collected from Baystate Medical Center, Springfield, Massachusetts and some reference books borrowed from library to support this project development. Besides that, the past research is vital as to provide information, guideline and better understanding of current developing project.