

**HYBRID METHOD MECHANISM OF
MULTI-LAYER FEED-FORWARD NEURAL
NETWORK (MLFFNN) FOR BIOMETRY
REGRESSION MODEL**

MOHAMAD NASARUDIN BIN ADNAN

UNIVERSITI SAINS MALAYSIA

2026

HYBRID METHOD MECHANISM OF MULTI-LAYER FEED-FORWARD NEURAL NETWORK (MLFFNN) FOR BIOMETRY REGRESSION MODEL

By

MOHAMAD NASARUDIN BIN ADNAN

Thesis submitted in fulfilment of the requirements

for the Degree of

Doctor of Philosophy

January 2026

ACKNOWLEDGEMENT

I begin by expressing my deepest gratitude to Allah for His health, protection, and guidance throughout my PhD journey. The path has not been without its challenges, but through divine blessing, I have been able to navigate it with relative ease and success.

I would like to extend my sincerest thanks to my main supervisor, Associate Professor Ts. Dr. Wan Muhamad Amir Bin W Ahmad, for his unwavering support and mentorship. His guidance, encouragement, patience, and invaluable insights in biostatistics have been crucial in shaping my academic growth. I am also grateful for the opportunities he provided to attend biostatistics workshops, which have enriched my learning experience, both before and during my PhD studies.

I would also like to express my heartfelt appreciation to my co-supervisor, Dr. Ruhaya Binti Hasan, for her guidance, constructive feedback, and continuous support throughout this research. Her insights and encouragement have contributed significantly to the completion of this thesis.

To my beloved family, I dedicate this thesis with heartfelt appreciation. To my parents, thank you for your endless love, support, and encouragement. Your belief in me has been a constant source of strength. To my siblings, your love and continuous encouragement have given me the courage to face every challenge and have been a pillar of hope throughout this journey.

Finally, I would like to express my sincere gratitude to the Ministry of Higher Education Malaysia for the Fundamental Research Grant Scheme [Code: FRGS/1/2022/STG06/USM/02/10], and to Universiti Sains Malaysia (USM) for awarding me the GRA-Assist, which alleviated my financial burden and made it possible for me to complete my PhD studies.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	iv
LIST OF FIGURES	v
LIST OF SYMBOLS AND ABBREVIATIONS	vii
LIST OF APPENDICES	ix
ABSTRAK	x
ABSTRACT	xii
CHAPTER 1 INTRODUCTION	1
1.1 Preview of the Chapter	1
1.2 Background of the Study – A Review of Statistical Modelling	1
1.3 Research Motivation and Research Problem/Statement	2
1.4 The Rationale of the Study	4
1.5 Conceptual Framework of the Research Study	7
1.6 Research Hypothesis	10
1.7 Research Questions	10
1.8 Research Objectives	11
1.8.1 General Objectives	11
1.8.2 Specific Objectives	11
1.9 Scope and Methodology	11
1.10 The Contribution of the Study	12
1.11 Limitation of the Study	13
1.12 Organization of the Thesis	15
CHAPTER 2 LITERATURE REVIEW	17
2.1 Overview	17
2.2 History of Regression in Science Fields	17
2.3 Biometry Modelling using a Regression Model	19

2.4	Biometry Modelling using the Ordered Logistic Regression Model	21
2.5	Biometry Modelling using the Multiple Linear Regression Model	22
2.6	Biometry Modelling using the Binary Logistic Regression Model	24
2.7	Accessibility of Multi-Layer Feed-Forward Neural Network (MLFFNN).....	25
2.8	Biometry Modelling using Surface Plot Methodology	27
2.8.1	Surface Plot Analysis in Health Sciences	28
2.8.2	Contour Plot Analysis in Health Sciences Applications	30
2.9	Bootstrapping Approach	31
2.10	Application of Statistical Methods.....	33
2.10.1	Case Study I: Type 2 Diabetes among Dyslipidemia Patients	33
2.10.2	Case Study II: A Forensic Crime Case	35
2.10.3	Case Study III: Hypertension among Dyslipidemia Patients	38
2.11	Review of Methodology and Limitations	39
2.12	Concluding Remarks	42
CHAPTER 3 METHODOLOGY		45
3.1	Overview of Proposed Methodology	45
3.2	Study Design	45
3.3	Study Population	46
3.4	Study Period.....	47
3.6	References and Source of the Population.....	47
3.7	Ethical Consideration	48
3.8	Sampling Frame	48
3.9	Software Used in Research	48
3.10	Variables Used in Research	48
3.11	The Main Components for the Methodology Building using the R syntax	50
3.11.1	Data Bootstrapping.....	51
3.11.2	Ordered Logistic Regression	51

3.11.3	Multiple Linear Regression	52
3.11.4	Binary Logistic Regression	54
3.11.5	Response Surface Plot (RSP)	56
3.11.6	Decision Tree Analysis	57
3.11.7	Multi-Layer Feed Forward Neural Network (MLFFNN)	60
3.12	The Proposed Method	64
3.12.1	R- Syntax for Hybrid Ordered Logistic Regression	64
3.12.2	R- Syntax for Integrated Multiple Linear Regression	72
3.12.3	R- Syntax for Integrated Multiple Logistic Regression	77
3.13	Summary	83
CHAPTER 4	RESULTS	84
4.1	Preview of Chapter.....	84
4.2	Case I: Integrated Ordered Logistic Regression: A Medical Application	84
4.3	Case Study I: Hypertension with Dyslipidemia and Type 2 Diabetes	85
4.3.1	Decision Tree Analysis and The Model Evaluation.....	87
4.3.3	Response Surface Plot (RSP)	90
4.4	Case II: Integrated Multiple Linear Regression: A Forensic Application	101
4.4.1	Case Study II: A Forensic Crime Case.....	101
4.4.2	Model Evaluation	103
4.4.3	Response Surface Plot (RSP)	104
4.5	Integrated Binary Logistics Regression: A Medical Application	122
4.5.1	Case Study III: Hypertension among Dyslipidemia Patients	122
4.6	Summary	136
CHAPTER 5	DISCUSSION.....	138
5.1	Chapter Overview	138
5.2	The Integrated Model Proposed	138
5.3	The Application of the Integrated Model.....	139

5.4	The Case Studies	141
5.4	Conclusion	142
CHAPTER 6 CONCLUSION		144
6.1	Chapter Overview	144
6.2	Summary and Conclusion	144
6.4	Recommendations and Future Directions	147
6.5	Potential for Future Work	149
6.6	Overall Conclusion.....	149
REFERENCES.....		150
APPENDICES		

LIST OF TABLES

		Page
Table 3.1	Data description of the selected variables for Case I	49
Table 3.2	Data description of the selected variables for Case II	49
Table 3.3	Data description of the selected variables for Case III	50
Table 4.1	Result of multiple logistic regression by combining the bootstrap method training and testing dataset	85
Table 4.2	Result of the Exp(B)	86
Table 4.3	Confusion Matrix and Statistics Reference	89
Table 4.4	Overall Statistics for Decision Tree Analysis	89
Table 4.5	The Summary of Statistics by Class	90
Table 4.6	Result of multiple linear regression by combining the bootstrap method training and testing dataset	103
Table 4.7	The “Actual” and “Predicted” values obtained through the proposed methodology	103
Table 4.8	Comparison of the “Actual Data” with “Predicted Data”	104
Table 4.9	Result of multiple logistic regression by combining the bootstrap method training and testing dataset	123
Table 4.10	Result of the Exp(B)	123

LIST OF FIGURES

		Page
Figure 1.1	Conceptual framework of the study- Ordered Logistic Regression Model	7
Figure 1.2	Conceptual framework of the study- Multiple Linear Regression Model	8
Figure 1.3	Conceptual framework of the study- Binary Logistic Regression Model	9
Figure 3.1	Decision trees are graphical models for describing sequential decision problems.	59
Figure 3.2	The general architecture of the MLFFNN is with one hidden layer, N input nodes, H hidden nodes, and one output node.	61
Figure 3.3	The general architecture of the MLFFNN with two hidden layers, N input nodes, and one output node (Sources)	62
Figure 3.4	The general architecture of the MLFFNN with one hidden layer, N input nodes, two H hidden nodes, and three output nodes.	64
Figure 3.5	Flowchart of the proposed statistically Ordered Logistic Regression Model	71
Figure 3.6	Flowchart of the proposed statistical Multiple Linear Regression Model	76
Figure 3.7	Flowchart of the proposed statistical Multiple Logistic Regression Model	82
Figure 4.1	The best (MLFFNN) model's architecture has six input variables, two hidden layers, and three output nodes.	87
Figure 4.2	The classification results from DTA analysis.	88
Figure 4.3	The surface plot for the Diabetic status with the selected variables	91
Figure 4.4	The contour plot for the Diabetic status with the selected variables	92
Figure 4.5	The surface plot for the number of times respondents	106

	was victimized in a year with the selected variables	
Figure 4.6	The contour plot for the Diabetic status with the selected variables	107
Figure 4.7	The best (MLFFNN) model's architecture has five input variables, one hidden layer, and one output node.	124
Figure 4.8	The surface plot for the Hypertension status with the selected variables	125
Figure 4.9	The contour plot for the Hypertension status with the selected variables	126

LIST OF SYMBOLS AND ABBREVIATIONS

ALRM	Applied Linear Regression Model
ANN	Artificial Neural Network
ANN	Artificial Neural Networks
BLR	Binary Logistic Regression
BMI	Body Mass Index
CI	Confidence Interval
COVID-19	Coronavirus Disease 2019
Ctree	Conditional Inference Trees
DFA	Discriminant Function Analysis
DTA	Decision Tree Analysis
F1X	Mental And Behavioral Disorders
F2X	Schizophrenia And Delusional Disorders
F4X	Somatoform Disorders
F6X	Adult Personality Disorders
HDL	High-Density Lipoprotein
HPV	Human Papillomavirus
KRK	Klinik Rawatan Keluarga
MAD	Mean Absolute Deviance
MLE	Maximum Likelihood Estimator
MLFFNN	Multilayer Feedforward Neural Network
MLogisticR	Multiple Logistic Regression
MLP	Multilayer Perceptron
MLPNN	Multi-Layer Perceptron Neural Network
MLR	Multiple Linear Regression

MSE	Mean Squared Error
MSE.net	Mean Square Error
Neg Pred Value	Negative Predictive Value
NN	Neural Networks
OLR	Ordered Logistic Regression
OSCC	Oral Squamous Cell Carcinoma
PMSE	Predicted Mean Squared Error
Pos Pred Value	Positive Predictive Value
RSM	Response Surface Methodology
RSP	Response Surface Plot
T2DM	Type 2 Diabetes Mellitus
WHO	World Health Organization

LIST OF APPENDICES

Appendix A	Ethical Approval Letter
Appendix B	R Software Syntax
Appendix C	List of Research Publications
Appendix D	Personal Biodata

MEKANISMA METODOLOGI HIBRID RANGKAIAN NEURAL HADAPAN SUAPAN BERBILANG LAPISAN (MLFFNN) UNTUK MODEL REGRESI BIOMETRI

ABSTRAK

Pemodelan statistik memainkan peranan yang amat penting dalam sains perubatan dan forensik, di mana keupayaan meramalkan data biologi dan penyiasatan yang kompleks secara tepat adalah kritikal. Model regresi tradisional telah banyak menyumbang dalam analisis data, namun sering gagal merungkai hubungan tidak linear dan corak berdimensi tinggi yang wujud dalam set data moden. Walaupun terdapat kemajuan dalam teknik statistik dan pembelajaran mesin, pendekatan sedia ada masih terhad dari segi ketepatan ramalan dan keteguhan, khususnya apabila digunakan pada data biometri yang kompleks seperti ramalan penyakit kronik dan analisis forensik. Wujud jurang metodologi dalam mengintegrasikan regresi konvensional dengan pembelajaran mesin bagi menghasilkan rangka kerja pemodelan yang bersatu, boleh dipercayai dan mudah ditafsir. Kajian ini bertujuan membangunkan dan menjelaskan mekanisme metodologi hibrid yang mengintegrasikan Regresi Logistik Tertib (OLR), Regresi Linear Berganda (MLR), dan Regresi Logistik Binari (BLR) bersama Rangkaian Neural Suapan Hadapan Berbilang Lapisan (MLFFNN), Plot Permukaan Tindak Balas (RSP), dan Analisis Pokok Keputusan (DTA) bagi pemodelan biometri. Model hibrid dibangunkan dengan menggabungkan teknik regresi klasik bersama MLFFNN untuk pengesanan corak tidak linear, disokong oleh kaedah pensampelan semula yang meluas bagi menstabilkan anggaran parameter. Data telah dipra-proses untuk menyingkirkan nilai luar biasa, diresampel sebanyak 1000 kali bagi menstabilkan anggaran, dan

dibahagikan kepada set latihan dan ujian. OLR digunakan untuk hasil ordinal, MLR bagi hasil berterusan, dan BLR bagi hasil binari, manakala setiap model regresi dipertingkatkan lagi menggunakan MLFFNN dengan penalaan lapisan tersembunyi, fungsi pengaktifan, dan kadar pembelajaran. RSP digunakan untuk memvisualisasikan interaksi pelbagai pemboleh ubah dan mengenal pasti tahap ramalan yang optimum, manakala DTA menghasilkan laluan keputusan berasaskan peraturan yang mudah ditafsir. Prestasi model dinilai menggunakan ukuran seperti ketepatan, sensitiviti, spesifisiti, Min Ralat Kuasa Dua (MSE), dan Min Ralat Sisihan Mutlak (MAD), serta dibandingkan dengan model kaedah tunggal. Rangka kerja hibrid yang dicadangkan ini secara konsisten mengatasi kaedah konvensional dalam tiga kajian kes iaitu: (i) diabetes jenis 2 dalam kalangan pesakit dislipidemia, (ii) kes jenayah forensik, dan (iii) hipertensi dalam kalangan pesakit dislipidemia. Model ini mencapai ketepatan ramalan yang lebih tinggi, berjaya mengurangkan MSE dan MAD dengan ketara, serta meningkatkan kefahaman melalui visualisasi DTA dan RSP. Dapatan ini membuktikan keupayaan model untuk menangkap corak tidak linear yang kompleks, memperkukuh kebolehpercayaan diagnostik, dan meningkatkan prestasi ramalan dalam konteks perubatan dan forensik. Kajian ini memperkenalkan satu mekanisme hibrid statistik-pembelajaran mesin yang telah disahkan, yang memajukan pemodelan regresi biometri serta menawarkan inovasi metodologi dan alat praktikal untuk diagnostik kesihatan dan penyiasatan forensik.

HYBRID METHOD MECHANISM OF MULTI-LAYER FEED-FORWARD NEURAL NETWORK (MLFFNN) FOR BIOMETRY REGRESSION MODEL

ABSTRACT

Statistical modeling plays a critical role in medical and forensic sciences, where reliable prediction of complex biological and investigative data is essential. Traditional regression models have contributed significantly but often fail to capture nonlinear relationships and higher-dimensional patterns that arise in modern datasets. Despite advances in statistical and machine learning techniques, existing approaches remain limited in predictive accuracy and robustness, particularly when applied to challenging biometry data such as chronic disease prediction and forensic analysis. A methodological gap persists in integrating conventional regression with machine learning to create a unified, reliable, and interpretable modeling framework. This study aims to develop and elucidate a hybrid methodological mechanism that integrates Ordered Logistic Regression (OLR), Multiple Linear Regression (MLR), and Binary Logistic Regression (BLR) with Multi-Layer Feed-Forward Neural Networks (MLFFNN), Response Surface Plot (RSP), and Decision Tree Analysis (DTA) for biometry modeling. A hybrid model was constructed by combining classical regression techniques with MLFFNN for nonlinear pattern detection, enhanced by extensive bootstrapping for robust parameter estimation. Data were preprocessed to remove outliers, resampled 1000 times to stabilize estimates, and divided into training and testing sets. OLR was applied for ordinal outcomes, MLR for continuous outcomes, and BLR for binary outcomes, with each regression output further optimized using MLFFNN by tuning hidden layers, activation functions, and learning rates. RSP was used to visualize multi-variable interactions and determine optimal

predictor levels, while DTA generated interpretable, rule-based decision paths. Model performance was assessed using accuracy, sensitivity, specificity, Mean Squared Error (MSE), and Mean Absolute Deviance (MAD), and benchmarked against single-method models. The proposed hybrid framework consistently outperformed conventional techniques across three applied case studies: (i) type 2 diabetes among dyslipidemia patients, (ii) forensic crime case, and (iii) hypertension among dyslipidemia patients. It achieved higher predictive accuracy, significantly reduced MSE and MAD, and provided improved interpretability through DTA and RSP visualizations. The findings demonstrate the model's ability to capture complex nonlinear patterns, strengthen diagnostic reliability, and enhance predictive performance across both medical and forensic domains. This research introduces a validated hybrid statistical-machine learning mechanism that advances biometry regression modeling, offering methodological innovation and practical tools for healthcare diagnostics and forensic investigations.

CHAPTER 1

INTRODUCTION

1.1 Preview of the Chapter

In this chapter, the thesis is comprehensively summarized. It begins by providing a clear discussion of the study's background. Subsequently, the following section addresses the research motivation and problem, followed by an exploration of the study's rationale. The conceptual framework is presented, offering a visual representation of the entire research process. Moving forward, the research hypothesis is elaborated upon in the subsequent section, while the objectives section of the study outlines both general and specific goals. Throughout this chapter, the focus remains on aspects such as the study's scope, methodology, and contribution, all while aligning with the research objectives. Towards the end of this chapter, the study's limitations are examined in the final section, followed by an explanation of the organization of the thesis, providing a detailed account of how the current analysis was conducted.

1.2 Background of the Study – A Review of Statistical Modelling

Statistical methods have become indispensable tools in contemporary scientific research, underpinning evidence-based inquiry across a wide range of disciplines, including economics, education, business management, and biological sciences. Within the biological domain, these methods are collectively referred to as biostatistics, a specialized branch of applied statistics concerned with the analysis, interpretation, and inference of data arising from biological and medical studies. The growing complexity of scientific data has further emphasized the need for robust statistical frameworks capable of elucidating relationships among variables and

supporting accurate prediction and decision-making.

Among the various statistical techniques, regression analysis occupies a central role in modeling and quantifying relationships between dependent and explanatory variables. Traditionally, regression methods have been widely adopted due to their interpretability, theoretical foundation, and effectiveness in estimating population-level effects. However, as modern biomedical and forensic datasets increasingly exhibit nonlinear patterns, high dimensionality, and complex interactions, conventional regression approaches may be insufficient when applied in isolation. This limitation has motivated researchers to explore hybrid methodologies that integrate classical regression models with advanced machine learning techniques.

In response to these challenges, the present study investigates the integration of traditional regression models with Multi-Layer Feedforward Neural Networks and complementary analytical tools to enhance predictive accuracy and model robustness. By combining the strengths of statistical interpretability and machine learning flexibility, this research aims to provide a comprehensive framework suitable for medical and forensic applications. The study contributes to the evolving body of knowledge on hybrid modeling strategies, offering methodological insights that support improved prediction, validation, and practical applicability in complex real-world datasets.

1.3 Research Motivation and Research Problem/Statement

Statistical modeling has long been an important part of both medical and forensic sciences, especially when it comes to figuring out how different factors are related to each other and to results. However, these relationships are often influenced by latent interaction effects among predictors and by systematic biases arising from

model assumptions, sampling limitations, and variable interdependence. Even while both sectors have made progress, it is still very hard to effectively model and anticipate complicated events, including chronic diseases and forensic investigations, particularly when conventional statistical models fail to adequately capture nonlinear interactions, effect modification, and bias propagation across correlated variables. Choosing the right and strong biometric models is very important for getting accurate and reliable results, as inappropriate model selection may introduce estimation bias, obscure meaningful interaction structures, and lead to misleading inferences. However, there is a gap in the use of modern statistical methods that could help fill in the gaps that are already present, especially methods capable of mitigating bias, modeling higher-order interactions, and improving robustness under complex data structures (Ahmad et al., 2020; Adnan et al., 2024).

This research aims to address this gap by introducing a novel hybrid statistical model that combines OLR, MLR, and BLR with MLFFNN, RSP, and DTA to establish the relationship between multiple independent and dependent variables. By design, this hybrid framework seeks to reduce model-specific bias and explicitly accommodate interaction effects that are difficult to represent within single-method approaches. Traditional methods, which often rely on a single modeling approach, are insufficient for capturing the complexity of health and forensic data, as they typically assume linearity, independence, or proportional effects that may not be held in real-world biomedical and forensic datasets. Therefore, the proposed hybrid model seeks to improve predictive accuracy by integrating the strengths of multiple methodologies, allowing interaction structures, nonlinear dependencies, and bias-adjusted estimates to be simultaneously explored, thereby enabling a more comprehensive and precise analysis of complex data.

The study will evaluate the newly developed approach using metrics such as Mean Squared Error (MSE), Mean Absolute Deviance (MAD), and accuracy measurements. These metrics will be used to assess the effectiveness of the model in predicting chronic disease patterns and forensic outcomes, particularly in the context of the Malaysian population. By combining multiple statistical methods, this research introduces an innovative approach to modeling that has remained largely unexplored in the Malaysian context, particularly in relation to chronic diseases and forensic investigations. Given the increasing complexity of these fields, a hybrid model is not only necessary but also holds the potential to provide more accurate, actionable insights.

1.4 The Rationale of the Study

In both medical and forensic sciences, the advancement of comprehensive and cohesive research procedures continues to be a significant problem, especially with the integration of diverse statistical tools. This research fills a major gap in the literature by introducing a hybrid methodological approach that combines OLR, MLR, and BLR with MLFFNN, RSP, and DTA. Figures 1.1 to 1.3 show how these strategies work together in a new way that makes it easier to understand and anticipate complex data, which is important in biometry and medical research.

In today's era, the demand for highly precise models with optimal outputs is rapidly growing. Despite significant advancements in statistical methods, many models suffer from limitations in their integration and practical application. This study aims to bridge this gap by combining high-accuracy models with high-reliability models, offering a more comprehensive and adaptable approach to data analysis. Moreover, a major deficiency in current research methodologies is the lack of a robust

model validation mechanism, especially in the context of biometry modeling. Validation techniques such as these are often overlooked, resulting in models that may lack generalizability and robustness when applied to new or unseen data. Addressing this deficiency is pivotal to ensuring that the developed models are not only accurate but also reliable and applicable in real-world scenarios.

The integration of established methodologies and the enhancement of existing techniques is a crucial step toward overcoming these challenges. By adapting and combining conventional statistical approaches with advanced machine learning algorithms, this study proposes a novel framework that addresses these gaps, enhancing the accuracy, reliability, and applicability of biometry regression models. Through this work, the study aspires to contribute meaningful solutions to the fields of medical diagnostics and forensic analysis, offering a more efficient and comprehensive approach to model development and validation.

Figures 1.1 to 1.3 show the conceptual framework that this study used to put together a complete method by merging OLR, MLR, and BLR with MLFFNN, RSP, and DTA for medical health cases and forensic sciences. Figure 1.1 illustrates the first strategy of the integration of sophisticated statistical and machine learning methodologies to create a hybrid model for the prediction and analysis of complicated data. It starts with gathering and preparing data, and then it moves on to creating MLFFNN, OLR, Response Surface Methodology (RSM), and DTA. The purpose of these models is to validate relationships between variables, predict outcomes, and visualize how different factors interact. MLFFNN is trained with 70% of the data and tested on 30%, while OLR analyzes ordered categories, RSM helps visualize variable interactions, and DTA identifies key factors influencing outcomes. The models are then applied to medical data to assess their performance. The final results are analyzed,

and the findings are compiled into a thesis for submission, contributing to a deeper understanding of the hybrid methodology's effectiveness in biometry regression modeling.

Figure 1.2 shows the second strategy employed in this work, which is to create a hybrid model that incorporates MLFFNN, MLR, and RSM. The first step is to gather and prepare the data. Then, 70% of the data is used for training and 30% for testing. This stage is to make the MLFFNN activation function. The next step is to make the MLR model and RSM. The purpose of this method is to check the connection between dependent and independent variables, make accurate predictions, and present the data in a way that makes it easier to draw statistical conclusions. After that, the model is employed on forensic data.

Figure 1.3 shows the study's third strategy, which is to build a binary logistic regression (BLR) model that works with MLFFNN and RSM. First, data is gathered, outliers are filtered out, and the dataset is divided into 70% training data and 30% testing data. First, the MLFFNN activation function is created. Then, the Logistic Regression Model is built, and RSM is used. The major purpose of this strategy is to check the link between dependent and independent variables, make sure predictions are correct, and use graphs to make statistical inferences easier. This methodology is utilized on medical data analysis. The final results from these case studies are analyzed, and the findings are incorporated into the thesis for submission, providing insights and conclusions based on the developed model.

1.5 Conceptual Framework of the Research Study

First Methodology

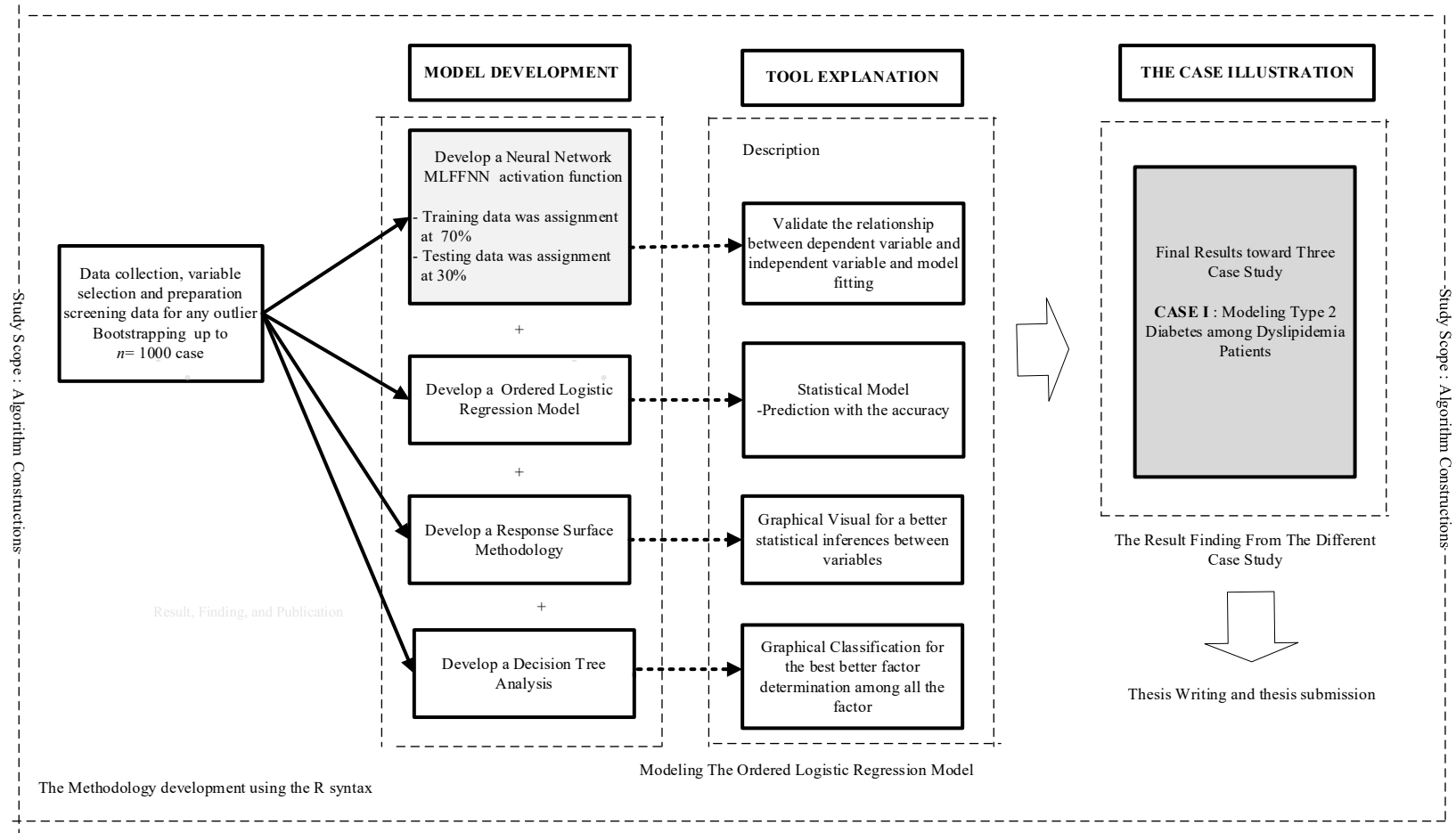


Figure 1.1: Conceptual framework of the study – Ordered Logistic Regression Model

Second Methodology

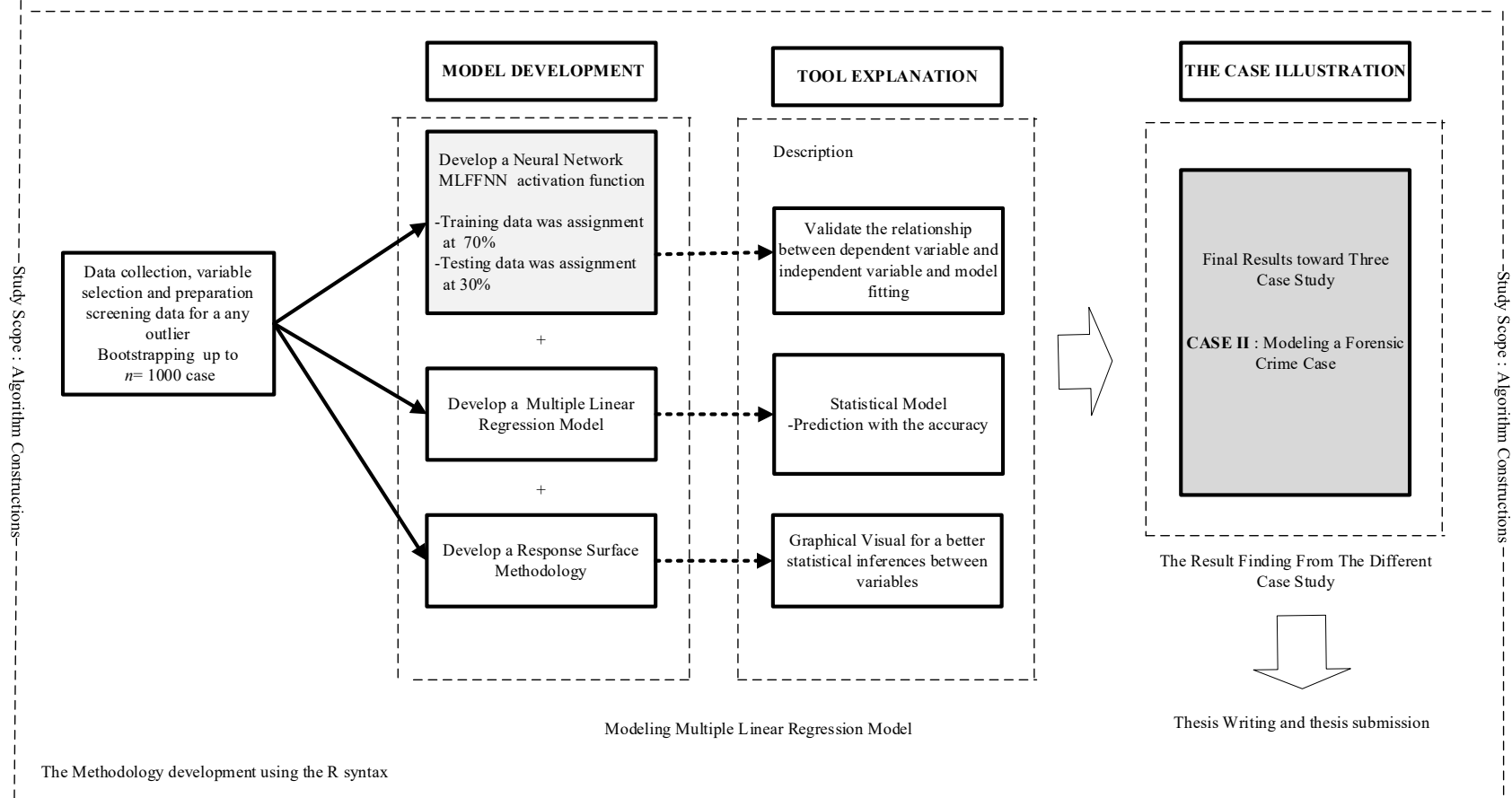


Figure 1.2: Conceptual framework of the study – Multiple Linear Regression Model

Third Methodology

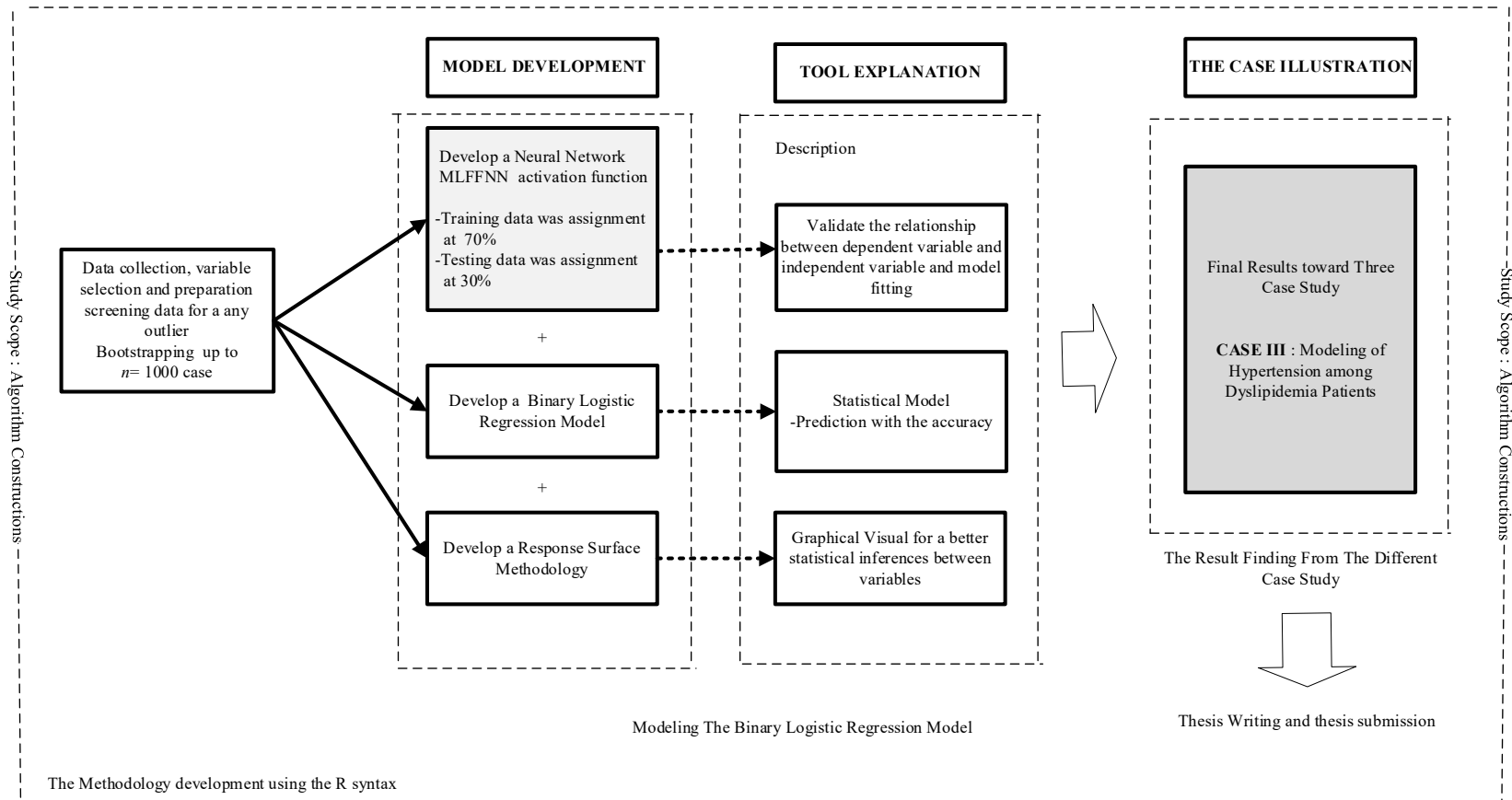


Figure 1.3: Conceptual framework of the study – Binary Logistic Regression Model

1.6 Research Hypothesis

This study is crucial to the development of foundational methods. The following hypotheses are stated.

H1: The proposed hybrid model will demonstrate significantly greater effectiveness and efficiency in enhancing the degree of analysis compared to existing conventional approaches.

H2: The hybrid model will achieve performance levels that allow it to be benchmarked as a standard framework for biometry modeling, reflecting improvements in precision, robustness, and applicability across datasets.

H3: The developed hybrid model will exhibit statistically significant predictive accuracy and reliability in identifying and validating chronic disease patterns, thereby supporting its clinical and epidemiological utility.

1.7 Research Questions

The following research questions are consequently presented to fulfill the study's objectives:

- a) How effective and efficient is the proposed hybrid model in enhancing the degree of analysis compared to conventional non-hybrid approaches?
- b) To what extent can the developed hybrid model be benchmarked as a standard framework for biometry modeling in comparison with existing statistical and computational methods?
- c) How accurately does the developed hybrid model perform in validating and predicting chronic disease patterns across diverse datasets?

1.8 Research Objectives

There are two categories of research objectives in this study: general objectives and specific objectives.

1.8.1 General Objectives

To develop and elucidate the hybrid method mechanism considering the Ordered Logistic Regression (OLR), Multiple Linear Regression (MLR), and Binary Logistic Regression (BLR) with Multi-Layer Feed-Forward Neural Network (MLFFNN), Response Surface Plot (RSP), and Decision Tree Analysis (DTA) for biometry modeling.

1.8.2 Specific Objectives

- a) To assess the effectiveness and efficiency of a hybrid model designed to enhance the depth and precision of analytical processes.
- b) To establish the hybrid model as a benchmark framework for biometry modelling.
- c) To validate the predictive performance of the hybrid model in identifying and forecasting chronic disease patterns.

1.9 Scope and Methodology

The scope of the study is to perform biometry modeling while integrating other statistical procedures. In this thesis, a hybrid method (known as the integrated method) has been constructed and introduced, which involves a combination of several main statistical methods that focus on more accurate research findings. The method proposed in this study revolves around OLR, MLR, and BLR with MLFFNN, RSP,

and DTA approach, as illustrated in Figure 1.1 - Figure 1.3. The method that has been developed will be implemented in three distinct case studies encompassing the disciplines of both medical and forensic sciences. The assessment of the three selected cases will encompass an extensive evaluation grounded in the outcomes obtained. This evaluation will primarily concentrate on the scrutiny of the accuracy and predictability of these cases, thereby affording an in-depth analysis of their efficacy in attaining the desired objectives.

1.10 The Contribution of the Study

The contribution can be summarized into two sections. The first contribution focuses on developing an integrated methodology that can be applied in medical and forensic crime cases for future predictions. This research endeavor is poised to make a substantial and impactful contribution to the evolution and refinement of methodological approaches, encompassing the utilization of the OLR, MLR, and BLR with MLFFNN, RSP, and the application of the DTA approach. By delving into these methodologies and their interplay, this study aims to foster a deeper understanding and exploration of their potential in diverse academic and practical domains. Through rigorous investigation and insights, this research aspires to broaden the knowledge, ultimately enriching the statistical tools for research practitioners. The second noteworthy contribution of this research pertains to the refinement and enhancement of the R-syntax framework for statistical tools.

This research enables researchers to customize the R-syntax to their individual needs, enhancing the performance and efficacy of the findings. In addition, this extended syntax provides a comprehensive and detailed output that caters to the specific requirements of researchers in terms of generating well-informed findings,

especially in the context of decision-making processes. This innovation in the R-syntax framework is a fantastic starting point for academics seeking statistical tools with high potential. Thirdly, this effort adds to the scientific literature and permits other researchers to continue future research work. This will aid the novice researcher in comprehending methodological challenges when designing new methods or conducting research. In addition, this study suggests an improved strategy for modeling purposes that is highly precise and predictable.

1.11 Limitation of the Study

This study has discerned two distinct constraints. The primary constraint emanates from methodological considerations, while the secondary limitation is the characteristics of the dataset and the software employed for analytical purposes. As a first limitation, only five parts of statistical tools were considered, and these were

a) **Bootstrapping**

Only involving case resampling of data, which aims to refine the parameters estimated.

b) **Ordered Logistic Regression**

For the first developed method (dependent variables as ordinal data): By exclusively employing the “enter” method, it is ensured that all the incorporated variables are already preprocessed and possess clinical significance, as evidenced in prior literature.

c) **Multiple Linear Regression**

For the second developed method (dependent variable as scale data): By exclusively employing the “enter” method, it is ensured that all the incorporated variables are already preprocessed and possess clinical significance, as

evidenced in prior literature.

d) Binary Logistic Regression

For the third developed method (dependent variable as binary data): By exclusively employing the “enter” method, it is ensured that all the incorporated variables are already preprocessed and possess clinical significance as evidenced in prior literature.

e) Multi-Layer Feed-Forward Neural Network

This methodology was used only to validate the factors obtained in (b), (c), and (d); calculate the accuracy and assess the methodology’s predictability. The predicted results (testing data) will be compared to the actual value, reflecting the proposed methods’ goodness.

f) Decision Tree Analysis

This methodology was employed to ascertain which of the factors included in the model made the most substantial contribution. Subsequently, this particular factor will be deemed the most crucial predictor for the case under investigation.

The second limitation pertains to the selection of the software utilized for this research endeavor. It is important to note that all the devised methodologies and techniques will be exclusively executed within the R software environment. R is chosen for its user-friendly interface and cost-effectiveness, as it is an open-source and freely available software tool. This constraint underscores the deliberate choice of R as the sole software platform for ensuring consistency, accessibility, and affordability in the implementation of research methodologies and techniques.

The data itself was the third limitation. For better comprehension and visualization in this research, only secondary data were used. The first scenario involved type 2 diabetes among dyslipidemia patients, the second case involved

forensic crime case, and the third case involved hypertension among dyslipidemia patients. All data was extracted from textbooks.

1.12 Organization of the Thesis

This thesis consists of six chapters and is structured as follows. The first chapter provides an introduction, concentrating on the background of the study and the setting in which it is being conducted. This chapter also addresses the study's aims, justification, scope, methodology, significance, and limitations.

A literature review of the statistical methods that will be integrated is offered in the second chapter. This chapter also discusses and investigates the statistical methodologies previously utilized in medical health and forensic criminal instances. Besides that, this chapter will discuss those strategies and their limits.

The third chapter is a methodology containing detailed information on the procedure and statistical models, namely the integration of OLR, MLR, and BLR with MLFFNN, facilitated by RSM, and the application of the DTA. This chapter explains the study design, study site, study period, and case studies in brief. This section also includes a study flowchart based on the proposed statistical method.

The fourth and fifth chapters of this thesis are dedicated to conducting an exhaustive and meticulous analysis of the research findings, accompanied by a comprehensive discussion that delves deeply into the specifics of each individual case study. These chapters undertake a meticulous examination, thoroughly exploring and elucidating the analysis. The aim is to provide a rich and detailed understanding of the findings within the context of each case study. This thorough analysis and discussion pave the way for a comprehensive and holistic comprehension of the research outcomes.

In the concluding chapter, an elaborate exposition of the conclusions derived from the outcomes of the findings is provided. This chapter further incorporates a concise deliberation regarding the analysis and interpretation of the model construction, along with several recommendations intended to enhance future research efforts.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

This chapter provides a comprehensive review of various statistical methods employed in biometry modeling. Sections 2.2 through 2.4 introduce and discuss key regression techniques, including Logistic Regression, Multiple Logistic Regression, and Multiple Linear Regression (MLR), highlighting their applications and significance in scientific research. Section 2.5 examines the bootstrapping approach, detailing its methodology and relevance in enhancing the reliability of statistical models. The accessibility and utility of Multi-Layer Feed-Forward Neural Networks (MLFFNN) in scientific fields are explored in Section 2.6, showcasing their potential in predictive modeling. In Sections 2.7 and 2.8, three case studies are presented, illustrating the application of the previously discussed statistical methods in real-world scenarios. Furthermore, Section 2.8 emphasizes the need to integrate multiple statistical techniques to overcome the limitations inherent in individual approaches. The chapter concludes with a summary of key findings and insights in Section 2.9.

2.2 History of Regression in Science Fields

In 1877, Sir Francis Galton, a pioneering British scholar, introduced the concept of regression in his landmark paper “Typical Laws of Heredity” in England. His work primarily focused on exploring causality and understanding the relationships between various variables. Galton’s groundbreaking experiments, conducted in 1886, examined the growth rates of peas in relation to their seed sizes. Dividing them into seven groups based on seed size, he measured the diameters of 100 seeds and found that smaller seeds tended to grow at a faster rate compared to larger ones. Moreover,

Galton observed that the average size of the seeds regressed towards a normal range, shedding light on the relationship between seed size and growth rate. This phenomenon, which he termed “regression,” marked the first documented use of the concept of statistical regression, which would later become integral to statistical modeling and analysis (Dhakal, 2018).

In addition to his work with peas, Galton extended his research to study the relationship between parental and offspring heights. By analyzing data from 205 observations, he identified a pattern where children of shorter parents were, on average, slightly taller than their parents, whereas children of taller parents tended to be somewhat shorter. This phenomenon, which he initially referred to as “revert,” was later renamed “Regression Toward Mediocrity.” This concept played a pivotal role in the development of linear regression, a cornerstone of modern statistics that remains fundamental in scientific research today. Galton’s insights into familial height dynamics laid the foundation for the statistical principles that would later evolve into regression analysis (Suarez et al., 2017).

The early 20th century saw the development of linear regression as a formalized method, further refined by Karl Pearson, who introduced the correlation coefficient in 1896, facilitating the measurement of the strength and direction of linear relationships between variables. Regression analysis gained prominence as a tool for various applications, particularly in the social and biological sciences, where understanding variable relationships is crucial for decision-making and predictions. More recent studies, such as those by Smith et al., (2020) and Baihaqi (2018), highlight the continued evolution of regression techniques in modern applications, including machine learning, biometry, and epidemiology. These advancements emphasize the significance of regression models in addressing complex scientific challenges and

enhancing predictive accuracy in diverse fields.

As the field of statistics grew, the concept of regression expanded, encompassing a variety of models designed to address specific types of data and relationships. Today, non-linear regressions, multi-variable models, and hybrid methods, such as the use of machine learning techniques, continue to advance the application of regression in science, particularly in areas like biometry and healthcare (O'Brien & Silcox, 2024).

2.3 Biometry Modelling using a Regression Model

Biometry modelling has gained widespread recognition in the realm of biostatistics, providing critical insights into biological processes and outcomes. One approach that has been utilized in biometry is regression analysis, a powerful statistical technique used to predict and estimate relationships between variables. Ahmad et al., (2017) explored the application of exponential growth regression modelling using SAS software, presenting it as an alternative method in biostatistics. Their study demonstrated how regression models could effectively analyze biological data, offering valuable insights into population growth, disease progression, and other biological phenomena. By employing the exponential growth model, researchers can derive more accurate predictions, making this approach an essential tool for biometry (Ahmad et al., 2017).

The development and evolution of regression as a method in biometry can be traced back to early statistical work, as discussed by Dhakal (2018). Dhakal's review highlights the historical significance of regression in statistical analysis, particularly its initial application by Francis Galton in the 19th century. Galton's groundbreaking research on hereditary traits, specifically the relationship between parent and offspring

height, led to the concept of “regression toward the mean”. This fundamental principle of regression formed the foundation for modern biostatistical models, where regression analysis now serves as a cornerstone in analyzing biological data. Regression’s flexibility allows it to be adapted to various types of data, making it indispensable in fields such as genetics, epidemiology, and clinical research (Dhakal, 2018).

In recent advancements, biometry modelling has expanded beyond simple linear relationships to include more complex models that capture intricate biological patterns. Krashniak and Lamm (2021) examined Galton’s contributions and the enduring relevance of regression models in modern statistics, particularly in the study of biological traits. Their work highlights the importance of regression analysis in understanding variability and relationships in biological data, underscoring the continued relevance of biometry modeling as a vital tool for researchers. The integration of machine learning techniques with traditional regression methods has further enhanced the predictive accuracy and robustness of these models, making them increasingly relevant in advanced biometry research (Krashniak and Lamm, 2021).

The application of regression models is not limited to basic biological analysis; it has also been successfully employed in other domains, such as engineering. Patel et al., (2020) utilized regression analysis to model and predict surface roughness, demonstrating the versatility of regression in capturing complex relationships in engineering datasets. While their study focused on manufacturing processes, the underlying principles of regression are directly transferable to biometry, where similar models can be employed to predict health outcomes, disease severity, and other biological measures. This broad applicability reinforces the value of regression models in biometry, providing researchers with the ability to derive predictive insights from complex and multi-dimensional data, thereby enhancing the accuracy of their findings

and supporting more informed decision-making (Patel et al., 2020).

2.4 Biometry Modelling using the Ordered Logistic Regression Model

Ordered Logistic Regression (OLR) is a widely used statistical technique for modeling ordinal outcomes, where the dependent variable consists of ordered categories with a natural ranking, but the exact spacing between categories is unknown. This method is highly valuable in biometry, where many medical and health outcomes-such as disease severity, stages of a condition, or quality of life-are measured on an ordinal scale. OLR models the relationship between a set of predictor variables and the odds of being at or below a certain category of the dependent variable. The underlying assumption of the model is proportional odds, meaning that the effect of predictors is assumed to be constant across the different thresholds of the ordinal outcome (Ghazali et al., 2020).

OLR has been effectively applied in numerous biometry studies, particularly in health and medical research. For example, in a study on the severity of Coronavirus disease 2019 (COVID-19) in China, researchers utilized OLR to identify key determinants influencing illness severity, including age, comorbidities, immune markers, and socio-demographic factors. This analysis provided important insights that could guide targeted interventions and improve disease management strategies (Xu et al., 2020). Similarly, Adnan et al., (2024) adopted a hybrid approach combining OLR and Multi-Layer Feed-Forward Neural Networks (MLFFNN) to predict factors associated with hypertension and dyslipidemia. The study highlighted significant predictors, including age, body mass index, lipid profile, and lifestyle factors, underscoring the potential of integrating traditional statistical methods with advanced machine learning techniques to enhance predictive accuracy and model robustness

(Adnan et al., 2024).

The validity of OLR models depends on several key assumptions, particularly the proportional odds assumption, which states that the effect of the predictors is consistent across all outcome thresholds. This assumption can be tested using methods like the Brant test. Additionally, researchers must ensure that multicollinearity does not distort the results, and the sample size should be sufficient to yield reliable estimates. Diagnostic checks, including assessing the proportionality assumption and examining residuals, are crucial for validating the appropriateness of the OLR model for a given dataset. Advances in computational tools, such as the `polr` function in R, have made OLR more accessible, allowing researchers to easily implement the model and interpret its results (Ghazali et al., 2020; Xu et al., 2020; Selman et al., 2024).

Recent innovations in hybrid methodologies have enhanced the predictive power of OLR in biometry. For example, Adnan et al., (2024) developed a hybrid model that combined OLR with bootstrapping and MLFFNN to predict hypertension risk, showing improved performance over traditional methods. These hybrid approaches demonstrate the growing trend of integrating statistical and machine learning techniques to address complex health-related problems. With continued advancements in computational capabilities and the development of hybrid models, OLR remains an essential tool in biometry, providing valuable insights for medical research, clinical decision-making, and personalized healthcare strategies (Adnan et al., 2024).

2.5 Biometry Modelling using the Multiple Linear Regression Model

Biometric modeling using the Multiple Linear Regression (MLR) model is a powerful statistical tool for exploring relationships between multiple independent

variables and a dependent variable. It is widely used in fields like health sciences and population studies, where it helps to understand complex interactions within data. Ahmad et al., (2016) advanced the MLR approach by combining bootstrap, and robustness techniques. This integration enhances the model's accuracy and adaptability, particularly when handling outliers or uncertain data, which are common in real-world biometric studies (Ahmad et al., 2016).

In health sciences, Ahmad et al., (2020) employed a MLR model to approximate randomized block designs, demonstrating its capability to handle complex relationships in health-related data. This method is beneficial in scenarios where traditional MLR may fail to account for uncertainty or variability in the data, allowing for more reliable results in health research. Such applications are crucial for understanding various health conditions and predicting outcomes based on multiple influencing factors (Ahmad et al., 2020).

In demographic studies, MLR has also proven helpful in analyzing factors affecting population health. For example, Palupi and Rizki (2020) used MLR to investigate the factors influencing infant mortality rates in East Java. By applying MLR to socio-economic data, they were able to identify key determinants of infant mortality, providing policymakers and healthcare providers in the region with valuable insights. This study demonstrates the model's effectiveness in public health research, where understanding demographic patterns is essential for developing targeted interventions (Palupi & Rizki, 2020).

Overall, the MLR model's ability to incorporate multiple variables and analyze their combined effects makes it a valuable tool in both health and demographic research. By adapting the model to address the specific needs of different datasets, researchers can uncover insights that inform policy decisions and health strategies,

ultimately improving outcomes in these fields (Ghazali et al., 2020; Rubi et al., 2021).

2.6 Biometry Modelling using the Binary Logistic Regression Model

Biometry modelling is a critical aspect of data analysis, particularly when studying binary outcomes, such as the presence or absence of a particular characteristic or event. One of the most commonly used techniques in this field is Binary Logistic Regression (BLR). This statistical method models the relationship between a binary dependent variable and one or more independent variables, making it ideal for applications in medical and forensic research. Binary Logistic Regression is particularly useful when the outcome of interest is categorical, such as estimating sex based on biometric data. This approach allows researchers to understand how predictor variables, such as measurements or demographic factors, influence the probability of a binary outcome (Rani et al., 2022).

The strength of Binary Logistic Regression lies in its ability to provide odds ratios, which are easy to interpret and communicate. This makes it a valuable tool in applied biometry, particularly in forensic research, where predicting outcomes such as sex, age, or ancestry based on biometric measurements is crucial. Harris (2021) emphasizes that BLR's ability to handle both continuous and categorical predictors makes it flexible and applicable to a wide range of biometric datasets. Furthermore, the model does not assume a linear relationship between independent and dependent variables, which is a significant advantage in dealing with complex biological data (Harris, 2021).

Additionally, BLR is widely used in medical research to predict the likelihood of diseases or other health-related outcomes. Schober and Vetter (2021) highlight the advantages of using logistic regression in clinical settings, as it enables the