FINAL THESIS BOOK
CAPSTONE DESIGN

Advanced Genomic Profiling and Classification of Breast Cancer Types: Leveraging Machine Learning Techniques for Precision Diagnostics

Compiled by:
Aurellia Rasya Gunawan / 1101213214
Ghina Mufidah/ 1101213452
Shania Wardani/ 1101213427
Tia Hasna Humayra/ 1101213479

TELECOMMUNICATIONS ENGINEERING BACHELOR'S DEGREE
FACULTY OF ELECTRICAL ENGINEERING
TELKOM UNIVERSITY BANDUNG
2025

APPROVAL PAGE

CAPSTONE DESIGN BOOK

ADVANCED GENOMIC PROFILING AND CLASSIFICATION OF BREAST CANCER TYPES: LEVERAGING MACHINE LEARNING TECHNIQUES FOR PRECISION DIAGNOSTICS

Has been approved and ratified as part of the Capstone Design
Telecommunication Engineering Undergraduate Program
Faculty of Electrical Engineering
Telkom University
Bandung

Compiled By:

Aurellia Rasya Gunawan / 1101213214
Ghina Mufidah/ 1101213452
Shania Wardani/ 1101213427
Tia Hasna Humayra/ 1101213479

Bandung, 13 January 2025
Approving,

Supervisor	Co-Supervisor
Suryo Adhi Wibowo, S.T., M.T., Ph.D.	Dr. Koredianto Usman, S.T., M. Sc.
NIP. 10870003	NIP. 2750053

APPROVAL PAGE
CAPSTONE DESIGN BOOK

ADVANCED GENOMIC PROFILING AND CLASSIFICATION OF BREAST CANCER TYPES: LEVERAGING MACHINE LEARNING TECHNIQUES FOR PRECISION DIAGNOSTICS

Has been approved and ratified as part of the Capstone Design
Telecommunication Engineering Undergraduate Program
Faculty of Electrical Engineering
Telkom University
Bandung

Compiled By:

Aurellia Rasya Gunawan / 1101213214
Ghina Mufidah/ 1101213452
Shania Wardani/ 1101213427
Tia Hasna Humayra/ 1101213479

Bandung, 13 January 2025
Approving,

Supervisor	Co-Supervisor
Suryo Adhi Wibowo, S.T., M.T., Ph.D.	Dr. Koredianto Usman, S.T., M. Sc.
NIP. 10870003	NIP. 2750053

ORIGINALITY STATEMENT SHEET

I, the undersigned:

Name	: Aurellia Rasya Gunawan
Student ID	: 1101213214
Address	: Jl. Cipinang Kebembem No. 2, RT.13/RW.13, Pisangan Timur, Kec. Pulo Gadung, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta (13230)
Phone Number	: +62 852-1099-6596
Email	: aurelliarasya27@gmail.com

I therefore declare that this Capstone Design Book is my original work, created in collaboration with my Capstone Design group, titled:

ADVANCED GENOMIC PROFILING AND CLASSIFICATION OF BREAST CANCER TYPES: LEVERAGING MACHINE LEARNING TECHNIQUES FOR PRECISION DIAGNOSTICS

I am prepared to accept the risks and repercussions that may be imposed on me in the future if this work is found to violate academic honesty or scientific ethics or if there is evidence of its originality.

Bandung, 18 December 2024

Aurellia Rasya Gunawan
1101213214

ORIGINALITY STATEMENT SHEET

I, the undersigned:

Name : Ghina Mufidah

Student ID : 1101213452

Address : Perumahan Taman Bukirsari Estate Kav. 07,
Jln. Raya Bukirsari, Tulusrejo, Lowokwaru, Kota Malang, Jawa Timur, (65141)

Phone Number : +62 812-5254-9993

Email : mufidahg17@gmail.com

I therefore declare that this Capstone Design Book is my original work, created in collaboration with my Capstone Design group, titled:

ADVANCED GENOMIC PROFILING AND CLASSIFICATION OF BREAST CANCER TYPES: LEVERAGING MACHINE LEARNING TECHNIQUES FOR PRECISION DIAGNOSTICS

I am prepared to accept the risks and repercussions that may be imposed on me in the future if this work is found to violate academic honesty or scientific ethics or if there is evidence of its originality.

Bandung, 18 December 2024

Ghina Mufidah

1101213452

ORIGINALITY STATEMENT SHEET

I, the undersigned:

Name : Shania Wardani

Student ID : 1101213427

Address : Sijeruk, Ngaren, Ngadirejo, Temangggung

Phone Number : +62 822-2135-1308

Email : shaniawrdn22@gmail.com

I therefore declare that this Capstone Design Book is my original work, created in collaboration with my Capstone Design group, titled:

ADVANCED GENOMIC PROFILING AND CLASSIFICATION OF BREAST CANCER TYPES: LEVERAGING MACHINE LEARNING TECHNIQUES FOR PRECISION DIAGNOSTICS

I am prepared to accept the risks and repercussions that may be imposed on me in the future if this work is found to violate academic honesty or scientific ethics or if there is evidence of its originality.

Bandung, 18 December 2024

Shania Wardani

1101213427

ORIGINALITY STATEMENT SHEET

I, the undersigned:

Name : Tia Hasna Humayra

Student ID : 1101213479

Address : Jl. Gatot Subroto, kp kali ulu, Kec. Cikarang Utara, Kabupaten Bekasi

Phone Number : +62 813-8600-2400

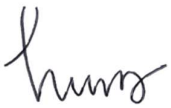
Email : tiahasna305@gmail.com

I therefore declare that this Capstone Design Book is my original work, created in collaboration with my Capstone Design group, titled:

ADVANCED GENOMIC PROFILING AND CLASSIFICATION OF BREAST CANCER TYPES: LEVERAGING MACHINE LEARNING TECHNIQUES FOR PRECISION DIAGNOSTICS

I am prepared to accept the risks and repercussions that may be imposed on me in the future if this work is found to violate academic honesty or scientific ethics or if there is evidence of its originality.

Bandung, 18 December 2024



Tia Hasna Humayra

1101213479

ABSTRACT

One of the biggest causes of death for women is still breast cancer and survival rates are significantly decreased by delayed identification. This study aims to employ machine learning approaches to classify breast cancer subtypes by utilizing sophisticated genomic profiling. The web-based application that was built makes use of data from genomic profiling. It provides precise and effective diagnostics for categorizing breast cancer by combining Artificial Neural Networks (ANN) with 30 features and with 6 features.

Rigorous testing of the models demonstrated their effectiveness in classifying breast cancer subtypes. The ANN model with 30 features achieved a remarkable accuracy rate of 99%, while the ANN model with 6 features achieved 100%, showcasing its superior ability to capture targeted genomic patterns with fewer features. The fundamental difference between the two models lies in the number of features used for training, where the ANN with 30 features incorporates a broader feature set. In contrast, the ANN with 6 features streamlines the analysis for efficiency. Additionally, the application features are optimized for each model: the ANN with 30 features provides comprehensive diagnostics, while the ANN with 6 features ensures a more targeted and simplified analysis.

The system's performance was validated through black box testing, confirming its reliability and usability in real-time scenarios. The application ensures seamless interaction and robust data handling by integrating Streamlit for an intuitive interface and Supabase for backend data storage. This platform offers healthcare providers a cost-effective and scalable solution for genomic analysis, facilitating early detection and personalized treatment strategies for breast cancer patients.

In conclusion, this research emphasizes the transformative potential of combining machine learning techniques for genomic profiling and cancer diagnostics. By bridging advanced computational models with real-world healthcare needs, the application contributes to developing innovative, accessible, and accurate solutions for improving breast cancer patient outcomes.

Keywords: Genomic Profile, Breast Cancer, Machine Learning, Classification, Website

PREFACE

First, the author would like to express the gratitude to the Almighty God for all His abundant grace and gifts so that the author can complete the thesis entitled "Advanced Genomic Profiling and Classification of Breast Cancer Types: Leveraging Machine Learning Techniques for Precision Diagnostics" well and on time. This thesis is compiled as one of the requirements to obtain a bachelor's degree and complete studies in the Telecommunication Engineering Bachelor's Degree program at Telkom University, Bandung.

The selection of this topic is motivated by the need for a faster and more accurate way to detect breast cancer. The previous method of detection is rather lengthy overall and may induce anxiety and panic in patients. This website is expected to classify breast cancer types effectively, quicken initial results, and assist medical personnel in clinical decision-making using machine learning-based technology in a broader commercial environment. This reduces the diagnosis time and gives patients reassurance and certainty sooner.

This thesis has limitations, and the author's knowledge in this field is still developing. Therefore, the author expects constructive input and suggestions to improve this work. Finally, the author would like to express deepest gratitude to all parties who have supported during the preparation of this thesis and apologize for any errors or shortcomings. The author hopes this research will be helpful for readers, academics, and practitioners in machine learning, and social media analysis.

Bandung. 18 December 2024

Author 1	Author 2	Author 3	Author 4
			
Aurellia Rasya Gunawan	Ghina Mufidah	Shania Wardani	Tia Hasna Humayra
1101213214	11001213452	1101213427	1101213479

ACKNOWLEDGMENT

With great humility and gratitude, the author wishes to express their heartfelt appreciation to all individuals and parties who have provided invaluable support, guidance, and encouragement throughout the completion of this thesis. Without their unwavering assistance, this research would not have been successfully completed.

First and foremost, the author would like to extend their deepest gratitude to:

1. Allah SWT, for His endless guidance, blessings, and strength that have illuminated the path and simplified every challenge throughout the research process.
2. Suryo Adhi Wibowo, S.T., M.T., Ph.D., as the first supervisor lecture, for his insightful guidance, expertise, and valuable advice that played a crucial role in the completion of this thesis.
3. Dr. Koredianto Usman, S.T., M.Sc., as the second supervisor lecture, for his comprehensive feedback and profound knowledge, which greatly enriched the depth and quality of this research.
4. Retno Hendriyanti, S.S., M.Pd., Ph.D., as the academic advisor, for her continued support, encouragement, and thoughtful advice throughout the course of this research.
5. Families, both immediate and extended, for their unwavering support, prayers, and motivation that served as a constant source of strength during the most challenging moments.
6. Classmates of TT45INT, for their camaraderie, collaboration, and shared journey that made this academic experience more fulfilling and memorable.
7. Friends in the Cyber Physical System Laboratory, for their shared knowledge, encouragement, and technical assistance that helped navigate through the complexities of this research.
8. Group mates of the DNA Group, for their teamwork, dedication, and collaborative spirit, which contributed significantly to the success of this project.
9. All staff of Telkom University, for their commitment, resources, and facilitation that created a conducive academic environment for learning and research.
10. All beloved individuals, for their continuous love, kindness, and support throughout the writing of this thesis, providing emotional strength and motivation to persevere until completion.

Lastly, the author is profoundly grateful to anyone whose names may not have been mentioned but whose contributions were equally meaningful. May all the assistance and support provided be repaid with blessings, and may this thesis serve as a valuable resource and inspire further research.

TABLE OF CONTENTS

APPROVAL PAGE i

CAPSTONE DESIGN BOOK i

ORIGINALITY STATEMENT SHEET ii

ORIGINALITY STATEMENT SHEET iii

ORIGINALITY STATEMENT SHEET iv

ORIGINALITY STATEMENT SHEET v

ABSTRACT vi

PREFACE vii

ACKNOWLEDGMENT viii

TABLE OF CONTENTS x

LIST OF FIGURES xiii

LIST OF TABLES xv

LIST OF ABBREVIATIONS 1

CHAPTER 1 PROPOSAL IDEAS 2

1.1 General Description of the Problem Background 2

1.2 Problem Analysis 2

1.2.1 Health Aspects 3

1.2.2 Economic Aspects 3

1.3 Capstone Objectives 3

1.4 Existing Solution Analysis 3

CHAPTER 2 LITERATURE REVIEW 5

2.1 Stable Connection for Distributed Database Access 6

2.2 Machine Learning 6

CHAPTER 3 DETAILED DESIGN OF THE SELECTED SOLUTION 8

3.1 Design of the Selection Solution 8

3.1.1 Intended Design System 8
3.1.2 Machine Learning 10
3.1.3 Website 11
3.2 Constraint and Specification 14
3.2.1 Constraint 14
3.2.2 Specification 14
3.3 Measurement or Verification Specifications 16
3.3.1 Patient's Dataset 16
3.3.2 Internet Connection 17
3.3.3 Result Visualization 18
3.3.4 Real-Time and Dataset Integration 19
3.3.5 Accuracy 19
3.3.6 Machine Learning 20
CHAPTER 4 IMPLEMENTATION 22
4.1 General Description of Implementation 22
4.2 Detail of The Implementation 22
4.2.1 Implementation of ANN 22
4.2.2 Implementation of Website 29
4.2.3 Implementation of Cloud Computing 32
4.3 Operating Procedures 38
CHAPTER 5 TESTING 43
5.1 General Testing Scenarios 43
5.1.1 SUS Testing Scenario 43
5.1.2 ANN Model Testing Scenarios 43
5.1.3 Model Deployment 44
5.1.4 Data Processing and Integration Testing Scenarios 44
5.1.5 Black Box Testing Scenario 45

5.2 Testing Details and Analysis 45
5.2.1 SUS 46
5.2.2 ANN Model with 30 Features Testing Details 48
5.2.3 ANN Model with 6 Features Testing Details 50
5.2.4 Model Deployment 53
5.2.5 Data Processing and Integration Testing Details 55
5.2.6 Black Box Testing Details 59
5.3 Operating Procedures 60
5.3.1 SUS Analysis Testing Result 60
5.3.2 ANN Model with 30 Features Testing Analysis 61
5.3.3 ANN Model with 6 features Testing Analysis 65
5.3.4 Model Deployment Testing Analysis 69
5.3.5 Data Processing and Integration Testing Analysis 69
5.3.6 Black Box Testing Analysis 69
CHAPTER 6 CONCLUSION AND SUGGESTIONS 70
6.1 Conclusion 70
6.2 Suggestions 71
BIBLIOGRAPHY 72
APPENDIX 75
APPENDIX A 76
SYSTEM USABILITY SCALE TESTING 76
APPENDIX B 77
BLACK BOX TESTING 77
APPENDIX C 81
BLACK BOX VISUAL EVIDENCE 81

LIST OF FIGURES

Figure 3. 1 Intended System Design 9
Figure 3. 2 ANN Workflow 10
Figure 3. 3 Website Flowchart 11
Figure 3. 4 Login Menu Flowchart 12
Figure 3. 5 Analysis Page Flowchart 13

Figure 4. 1 Website Implementation Diagram **22**
Figure 4. 2 Flowchart of ANN Implementations **23**
Figure 4. 3 ANN Architecture Model **28**
Figure 4. 6 Cloud Computing Workflow **32**
Figure 4. 7 Dashboard Website Breast Cancer Analysis **39**
Figure 4. 8 Login Form to Website of Breast Cancer Analysis **39**
Figure 4. 9 Sign Up to Website Breast Cancer Analysis **40**
Figure 4. 10 Warning Page on Analysis **41**
Figure 4. 11 Analysis Page **41**
Figure 4. 12 Column Input **42**
Figure 4. 13 Submit and Prediction Button **42**
Figure 4. 14 Logout Button **42**

Figure 5. 1 Classification Report of ANN with 30 Features Model 49
Figure 5. 2 Performance of The Model 50
Figure 5. 3 Hardware Characteristics 50
Figure 5. 4 Classification Report of ANN with 6 Features Model 52
Figure 5. 5 Performance of The Model 52
Figure 5. 6 PDF Format Result 55
Figure 5. 7 Data Processing and Integration 56
Figure 5. 8 Error Handling for Empty Field 58
Figure 5. 9 Result Prediction and Inserted Data to Database Notifications 58
Figure 5. 10 Data Tables from Database 59
Figure 5. 11 Data Security Using Secret 59
Figure 5. 12 SUS Scores Comparison 61
Figure 5. 13 ANN Model Confusion Matrices 62

Figure 5. 14 Confusion Matrices on Each Class 63
Figure 5. 15 Sensitivity and Specificity in 30 Features 65
Figure 5. 16 ANN with 6 Features Model Confusion Matrices 66
Figure 5. 17 Confusion Matrices on Each Class 66
Figure 5. 18 Sensitivity and Specificity in 6 Features 68

LIST OF TABLES

Table 3. 1 Product Specifications 15
Table 3. 2 Patient's Dataset 16
Table 3. 3 Internet Connection 17
Table 3. 4 Result Visualizations 18
Table 3. 5 Real-time and Dataset Integration 19
Table 3. 6 Accuracy 20
Table 3. 7 Machine Learning 21

Table 4. 1 Implementation of Website Application Interface 29
Table 4. 2 User Database Table 34
Table 4. 3 30 Features Database Table 34
Table 4. 4 6 Features Database Table 37

Table 5. 1 Google Collaboratory Specifications in Testing the Model 44
Table 5. 2 System Memory and Disk Usage Statistics . Error! Bookmark not defined.
Table 5. 3 Hardware Specifications in Black Box Testing 45
Table 5. 4 Results of 30 Features and 6 Features Trials 55
Table 5. 5 Percentile Range 60
Table 5. 6 Equation Value Results 63
Table 5. 7 Equation Value Results 67

LIST OF ABBREVIATIONS

ANN	: Artificial Neural Network
API	: Application Programming Interface
CPU	: Central Processing Unit
CSS	: Cascading Style Sheets
CSV	: Comma-Separated Values
FN	: False Negative
FP	: False Positive
GPU	: Graphics Processing Unit
HTML	: Hyper Text Markup Language
JSON	: Java Script Object Notation
JWT	: JSON Web Token
ML	: Machine Learning
PDF	: Portable Document Format
RAM	: Random Access Memory
ReLU	: Rectified Linear Unit
SQL	: Structured Query Language
SUS	: System Usability Scale
TN	: True Negative
TN	: True Negative
TP	: True Positive
TP	: True Positive
TPU	: Tensor Processing Unit
UI	: User Interface
URL	: Uniform Resource Locator
WHO	: World Health Organization