UNDERGRADUATE THESIS

COUGH CLASSIFICATION FOR EARLY DETECTION OF RESPIRATORY DISEASES USING A COMBINATION OF MEL-FREQUENCY CEPSTRAL COEFFICIENTS (MFCC) AND CAPSULE NETWORK

Desta Rizky Andhika
NPM 22081010072

THESIS ADVISORS
Dr. Faisal Muttaqin, S.Kom., M.T
Fetty Tri Anggraeny, S.Kom., M.Kom.

MINISTRY OF HIGHER EDUCATION, SCIENCE, AND TECHNOLOGY
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FACULTY OF COMPUTER SCIENCE
INFORMATICS STUDY PROGRAM
SURABAYA
2026

APPROVAL SHEET

COUGH CLASSIFICATION FOR EARLY DETECTION OF RESPIRATORY DISEASES USING A COMBINATION OF MEL-FREQUENCY CEPSTRAL COEFFICIENTS (MFCC) AND CAPSULE NETWORKS

By:

DESTA RIZKY ANDHIKA

NPM: 22081010072

Has been defended before, and accepted by, the Board of Assessors of the Thesis Examination of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, on August 20, 2026:

Approved.

Dr. Faisal Muttaqin, S.Kom., M.T. **NIP 19851231 2021211 009**	....	**(Advisor I)**
Fetty Tri Anggraeny, S.Kom., M.Kom **NIP.19820211 2021212 005**	....	**(Advisor II)**
Dr. Intan Yuniar Purbasari, S.Kom., MSc **NIP.19800602 2025212 029**	....	**(Lead Assessor)**
Hazna At Thooriqoh, S.Tr.Kom., M.Kom **NIP.19970320 202406 2 003**	....	**(Assessor I)**

Acknowledge by,
Dean of the Faculty of Computer Science

Prof. Dr. Ir. Novirina Hendrasarie, M.T.
NIP.19681126 199403 2 001

APPROVAL SHEET

COUGH CLASSIFICATION FOR EARLY DETECTION OF RESPIRATORY DISEASES USING A COMBINATION OF MEL-FREQUENCY CEPSTRAL COEFFICIENTS (MFCC) AND CAPSULE NETWORKS

By:

Desta Rizky Andhika

NPM: 22081010072

Approved to proceed to the Thesis Examination

Approved by,

Coordinator of Informatics Study Program
Faculty of Computer Science

Dr. Intan Yuniar Purbasari, S.Kom., MSc.
NIP.198006022 025212 029

STATEMENT OF ORIGINALITY

I, the undersigned:

Student Name/NPM	: Desta Rizky Andhika/ 22081010072
Degree Program	: Bachelor's (S1)
Major	: Informatics
Faculty	: Faculty of Computer Science

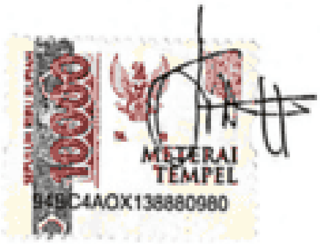
I hereby declare that this undergraduate thesis contains no part of any other scientific work that has been submitted to obtain an academic degree at any institution of higher education. Furthermore, it does not contain any work or opinions previously written or published by others, except for those that are explicitly cited in this thesis and listed in full in the references.

I further declare that this scientific document is free from any elements of plagiarism. Should any evidence of plagiarism be found in this thesis in the future, I am willing to accept sanctions in accordance with applicable laws and regulations.

I have made this statement of my own free will, without coercion from anyone, and it is to be used as intended.

Surabaya, September 1, 2026

Declarant,



Desta Rizky Andhika

NPM 22081010072

ABSTRACT

Student Name/NPM : Desta Rizky Andhika/22081010072

Thesis Title : Cough Classification for Early Detection of Respiratory Diseases Using a Combination of Mel-Frequency Cepstral Coefficients (MFCC) and Capsule Networks

Advisors 1. Dr. Faisal Muttaqin, S.Kom., M.T.

2. Fetty Tri Anggraeny, S.Kom., M.Kom.

Respiratory diseases such as acute respiratory infections (ARI), pneumonia, and chronic obstructive pulmonary disease (COPD) remain major global causes of morbidity, with cough serving as a crucial early clinical indicator. In conventional medical practice, distinguishing between dry (non-productive) and wet (productive) coughs is often subjective. This study aims to develop an artificial intelligence-based cough classification system as a non-invasive early screening tool for respiratory conditions using Mel-Frequency Cepstral Coefficients (MFCC) feature extraction combined with a Capsule Network (CapsNet) architecture. The methodology began by curating a public Kaggle dataset consisting of 27,550 raw metadata entries into a balanced, high-confidence experimental subset of 800 verified audio samples (400 dry and 400 wet). The preprocessing pipeline involved WAV conversion, mono reduction, 16 kHz resampling, silence trimming, adaptive RMS energy-based cough event segmentation, and duration standardization to 3 seconds (48,000 samples). Spectral features were extracted as MFCC coefficients and normalized using Z-scores. Model evaluation encompassed ten experimental scenarios (S01–S10), each run repeatedly across 10 runs with distinct random seeds (seeds 42–51) to assess model stability and performance. Experimental results demonstrated that scenario S02 (MFCC13-only, 2 routing iterations, class capsule dimension of 16, dropout rate of 0.40, and primary capsule stride of 3x8) achieved the highest and most stable average performance among all single and ensemble scenarios. Scenario S02 achieved a mean Balanced Accuracy of 71.67% ± 2.08%, an Accuracy of 71.67%, an F1-Score of 69.77%, a Precision of 71.00%, a Recall of 65.15%, a Specificity of 77.83%, a Macro-F1 of 71.45%, and a Matthews Correlation Coefficient (MCC) of 0.4384. The optimal model was subsequently deployed into a web-based screening prototype built with Streamlit. This research confirms that combining MFCC with Capsule Networks effectively captures the hierarchical spectral relationships of cough sounds to support AI-driven medical prescreening.

Keywords: Cough Classification,Early Detection of Respiratory Diseases, MFCC, Capsule Network, Dynamic Routing

ACKNOWLEDGMENTS

Praise and gratitude be to Allah SWT for all His blessings, mercy, and grace, which enabled the author to complete this undergraduate thesis titled **“Cough Classification for Early Detection of Respiratory Diseases Using a Combination of Mel-Frequency Cepstral Coefficients (MFCC) and Capsule Networks”** properly and on time. This thesis is submitted as one of the requirements to complete the Bachelor’s degree (S1) and earn the degree of Bachelor of Computer Science (S.Kom.) in the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” East Java.

In the process of preparing and completing this thesis, the author recognizes that many people have provided help, guidance, support, motivation, and prayers. Therefore, with all humility, the author would like to express sincere gratitude to:

1. The Rector of Universitas Pembangunan Nasional “Veteran” East Java, who provided the author with the opportunity to study and complete his education at Universitas Pembangunan Nasional “Veteran” East Java.
2. The Dean of the Faculty of Computer Science at the "Veteran" National Development University of East Java, who provided facilities and support throughout the author’s educational journey.
3. The Head of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" East Java, Dr. Intan Yuniar Purbasari, S.Kom., M.Sc., who provided facilities and support during the author’s studies.
4. Dr. Faisal Muttaqin, S.Kom., M.T., and Fetty Tri Anggraeny, S.Kom., M.Kom., as thesis advisors, who have dedicated their time, energy, and insights to provide guidance, direction, feedback, and motivation to the author throughout the process of writing this thesis. Their valuable suggestions and evaluations have greatly assisted the author in completing this research.
5. Dr. Intan Yuniar Purbasari, S.Kom., M.Sc., and Hazna At Thooriqoh, S.Tr.Kom., M.Kom., as thesis examiners, who have devoted their time, attention, constructive criticism, suggestions, and feedback to help improve this thesis.
6. All faculty members of the Informatics Study Program, who have shared their knowledge, experiences, and valuable lessons with the author throughout the author’s academic journey at the Faculty of Computer Science.
7. My beloved family, who have continuously offered prayers, love, support, motivation, and invaluable sacrifices during the author’s studies until the

completion of this thesis. All achievements obtained by the author are inseparable from their endless prayers and support.

8. A student from Surabaya State University who played a vital role in the author's journey. Thank you for being there, for your companionship, encouragement, for listening to my complaints, and for being a shoulder to lean on during the long process of completing this thesis. Thank you for your time, attention, prayers, and all the support you provided. May all the kindness we've shared remain a cherished memory, and may each of us find happiness and the best path in life.
9. Colleagues from HIMATIFA BLJ Dewaraksa—including Herald, Ghab, Feo, Fidel, Anjas, Theo, Henri, Dio, Satrio, and Evril—who were part of the author's journey during college. Thank you for the camaraderie, shared experiences, collaboration, support, and memories we created together. These experiences have become invaluable components of both my academic journey and personal growth.
10. Close friends from Pemuda Hijrah: Adib, Abdi, Daffa, Dani, Rehan, Bagas, Ipan, and Nantha, who have served as a source of shared stories, encouragement, support, and camaraderie throughout the author's university life until the completion of this thesis.

The author acknowledges that this thesis still has shortcomings and limitations, both in terms of research and writing. Therefore, the author warmly welcomes constructive criticism and suggestions for future research development.

Finally, the author hopes that this thesis will be of benefit to readers and contribute to the advancement of knowledge, particularly in the fields of audio signal processing, machine learning, and cough sound classification. May all the help, support, prayers, and kindness extended to the author be rewarded by Allah SWT.

Surabaya, September 1, 2026

Author

TABLE OF CONTENTS

ABSTRACT vii
ACKNOWLEDGMENTS ix
TABLE OF CONTENTS xi
LIST OF FIGURES i
LIST OF TABLES iii
CHAPTER 1 1
INTRODUCTION 1
1.1 Background 1
1.2 Problem Statement 4
1.3 Research Objectives 5
1.4 Research Objectives 5
1.5 Scope and Limitations 6
CHAPTER II 9
LITERATURE REVIEW 9
2.1 Previous Research 9
2.2 Theoretical Framework 21
2.2.1 Respiratory Diseases and the Importance of Early Detection 21
2.2.2 Cough as a Clinical Indicator 24
2.2.3 Cough Type Identification Based on Audio Analysis 27
2.2.4 Machine Learning in Healthcare 28
2.2.5 Deep Learning for Audio Classification 29
2.2.6 Audio Preprocessing in Signal Preprocessing 32
2.2.7 Mel-Frequency Cepstral Coefficients (MFCC) 40
2.2.8 Capsule Network (CapsNet) 56
2.2.9 Model Performance Evaluation 70
CHAPTER III 80
RESEARCH METHODOLOGY 80
3.1 Research Workflow 80

3.2 Literature Review and Problem Identification 81
3.3 Dataset Collection 82
3.4 Data Labeling and Selection 89
3.5 Audio Preprocessing 91
3.5.1 Raw Cough Audio Input 93
3.5.2 Audio Format Conversion to WAV 94
3.5.3 Stereo-to-Mono Conversion 95
3.5.4 Audio Resampling 96
3.5.5 Audio Trimming (Silence Removal) 97
3.5.6 Adaptive Energy-Based Cough Event Segmentation 98
3.5.7 Audio Padding (Duration Standardization) 99
3.6 Audio Quality Audit and Balanced-Confidence Dataset Formation 101
3.7 Data Splitting (Train-Validation-Test Split) 102
3.8 Feature Extraction Using MFCC 105
3.8.1 Step 1: Pre-Emphasis Filter 106
3.8.2 Step 2: Framing (Dividing the Audio Signal into Frames) 108
3.8.3 Step 3: Windowing (Hamming Window) 109
3.8.4 Step 4: Fourier Transform (Fast Fourier Transform/FFT per Frame) . 110
3.8.5 Step 5: Frequency Scaling to the Mel Scale (*Mel Frequency Scaling*) 112
3.8.6 Step 6: Mel Filterbank (Filter Energy Calculation) 113
3.8.7 Step 7: Logarithmic Compression (*Log Energy*) 114
3.8.8 Step 8: Discrete Cosine Transform 115
3.9 MFCC Feature Dataset and Labels 116
3.10 Capsule Network (CapsNet) Model Training 119
3.10.1 Step 1: Input Data Preparation 121
3.10.2 Step 2: Initial Feature Extraction Using a Convolutional Layer 122
3.10.3 Step 3: Primary Capsules Formation 124
3.10.4 Step 4: Vector Normalization Using the Squash Function 125
3.10.5 Step 5: Formation of Prediction Vectors Between Capsules *(Prediction Vectors*) 126
3.10.6 Step 6: Routing-by-Agreement Mechanism 127

3.10.7 Step 7: Determining Class Predictions 130
3.10.8 Step 8: Loss Function Calculation (Margin Loss) 131
3.10.9 Step 9: Backpropagation and Weight Updates 132
3.11 Model Training and Testing Scenarios 133
3.12 Model Performance Evaluation Method 136
3.13 Implementation of the Web-Based Screening System Prototype 140
CHAPTER IV 144
RESULTS AND DISCUSSION 144
4.1 Research Implementation Environment 144
4.2 Selection Results, Quality Audit, and Final Dataset Formation 147
4.3 Cough Audio Preprocessing and Segmentation Results 153
4.4 Results of Training, Validation, and Test Data Splitting 164
4.5 MFCC Feature Extraction and Representation Results 167
4.6 Implementation of the Capsule Network Architecture 173
4.7 Model Training Scenarios and Results 177
4.7.1 Scenario S01: Baseline MFCC13-Only 183
4.7.2 Scenario S02: Reduction of Routing Iterations 190
4.7.3 Scenario S03: Simplification of Class Capsule Dimensions 197
4.7.4 Scenario S04: Reduction of the Dropout Value 203
4.7.5 Scenario S05: Primary Capsule Stride Adjustment 210
4.7.6 Scenario S06: MFCC20-Only 216
4.7.7 Scenario S07: MFCC13-Only Soft-Voting Ensemble 223
4.7.8 Scenario S08: Weighted Soft-Voting Ensemble 226
4.7.9 Scenario S09: Weighted Soft-Voting Ensemble with Diverse Model Combinations 229
4.7.10 Scenario 10: Five-Model Ensemble Weighted Soft-Voting 231
4.8 Selection of the Best Model 234
4.9 Evaluation of the Best Model and Confusion Matrix Analysis 239
4.10 Implementation of a Web-Based Screening System Prototype 242
CHAPTER V 247
CONCLUSION AND RECOMMENDATIONS 247

5.1 Conclusion 247

5.2 Recommendations 248

REFERENCES 249

LIST OF FIGURES

Figure 2.1 Mechanism of Coughing 24
Figure 2.2 MFCC extraction workflow 40
Figure 2.3 Example of an audio signal before and after the pre-emphasis process with α = 0.95 42
Figure 2.4 Illustration of the 44
Figure 2.5 Illustration of 44
Figure 2.6 Illustration of 46
Figure 2.7 Illustration of a Mel Filterbank with triangular filters on the Mel frequency scale 48
Figure 2.8 Basic Architecture of CapsNet (Capsule Network) 57

Figure 3.1 Research Process Flowchart 80
Figure 3.2 Flowchart of Literature Review and Problem Identification 81
Figure 3.3 Flowchart of Dataset Collection 83
Figure 3.4 Kaggle Dataset Sources 83
Figure 3.5 Contents of the 84
Figure 3.6 Flowchart of the Data Labeling and Selection Process 89
Figure 3.7 Number of 91
Figure 3.8 Preprocessing Stage Flowchart 92
Figure 3.9 Example of 93
Figure 3.10 Example of Audio Converted to WAV Format 95
Figure 3.11 Flowchart of the Audio Quality Audit Process and Creation of a Balanced-Confidence Dataset 101
Figure 3.12 MFCC Flowchart 106
Figure 3.13 Flowchart of the CapsNet Model Training Process 121
Figure 3.14 Model Training and Testing Flowchart 134
Figure 3.15 Flowchart of the Web-Based Screening System Prototype Implementation 141
Figure 3.16 Input 142
Figure 3.17 Display of Screening Results and System Diagnostics 142

Figure 4.1 Examples of “Dry” and “Wet” Results in the Adaptive Energy Segmentation Process 161
Figure 4.2 Results of Adaptive Energy-Based Cough Event Segmentation 161
Figure 4.3 MFCC20 Heatmap for Training Sample 0 170
Figure 4.4 MFCC13 Heatmap for Training Sample 0 171
Figure 4.5 Comparison of Mean and Standard Deviation of Evaluation Metrics Across Scenarios in 10 Runs 180

Figure 4.6 Stability of 181
Figure 4.7 Distribution of Balanced Accuracy for Each Scenario Across 10 Runs 182
.Figure 4.8 Summary of the 188
Figure 4.9 Aggregate Confusion Matrix for S01 Across 10 Runs 189
Figure 4.10 Summary of the 195
Figure 4.11 Training Curve for S02, Run 2 (Best Run)..................................... 195
Figure 4.12 Aggregate Confusion Matrix for S02 Across 10 Runs 196
Figure 4.13 Summary of the 202
Figure 4.14 Aggregate Confusion Matrix for S03 Across 10 Runs 203
Figure 4.15 S04 Training Curve for 10 Runs... 208
Figure 4.16 Aggregate Confusion Matrix for S04 Across 10 Runs 209
Figure 4.17 Training Curve for S05 over 10 Runs 214
Figure 4. 18 Aggregate Confusion Matrix for S05 Across 10 Runs 215
Figure 4.19 Training Curve for S06 Over 10 Runs 221
Figure 4.20 Aggregate Confusion Matrix for S06 Across 10 Runs 222
Figure 4.21 Aggregate Confusion Matrix for S07 Across 10 Runs 225
Figure 4.22 Aggregate Confusion Matrix for S08 Across 10 Runs 227
Figure 4.23 Aggregate Confusion Matrix for S09 Across 10 Runs 230
Figure 4.24 Aggregate Confusion Matrix for S10 Across 10 Runs 233
Figure 4.25 Graph of 238
Figure 4.26 Dashboard View of the 243
Figure 4.27 Audio Input and Analysis Page... 244
Figure 4.28 Display of Prediction Results and Audio Quality Information 245

LIST OF TABLES

Table 2.1 Previous Research 13
Table 2.2 Common Causes of 25
Table 2.3 Symptoms Often Associated with a Dry Cough 25
Table 2.4 Sputum Characteristics in Productive Coughs and Possible Disorders . 27
Table 2.5 Advantages of the 52
Table 2.6 Advantages and Limitations of CapsNet 67
Table 2.7 Structure of the Confusion Matrix in Binary Classification 78

Table 3.1 Data Characteristics and Research Data Sources 84
Table 3.2 Data Distribution by Cough Type 88
Table 3.3 Structure of the 89
Table 3.4 Criteria for Audio Data Segmentation and Selection 90
Table 3.5 Example of Cough Audio Format Conversion to WAV Format 94
Table 3.6 Example of Manual Calculation of an Audio Recording into Frames .. 98
Table 3.7 Example of 99
Table 3.8 Example of 101
Table 3.9 Example of 102
Table 3.10 Balanced-Confident Dataset Split 105
Table 3.11 Example of Preprocessing Output Extraction Implementation 107
Table 3.12 Example of Spectrum Values and Filter Weights 114
Table 3.13 Structure of the 118
Table 3.14 Key Components in Capsule Network Training 119
Table 3.15 Input Data Structure for CapsNet Training (Input Preparation Stage) 122
Table 3.16 Structure of Primary Capsules in CapsNet Training 124
Table 3.17 Examples of Changes in Inter-Capsule Routing Coefficients 128
Table 3.18 Class Capsule Output Structure in the CapsNet Model 129
Table 3.19 Training Parameters Used in the Single CapsNet Scenario 135
Table 3.20 Single CapsNet Scenario 135
Table 3.21 Ensemble Scenario as a Comparison 136
Table 3.22 Summary of Model Performance Evaluation Metrics 137
Table 3.23 Example Confusion Matrix 138

Table 4.1 Specifications of the Research Implementation Environment 144
Table 4.2 Summary of Audio Data Label Selection Results 149
Table 4.3 Summary of 150
Table 4.4 Summary of Balanced-Confidence Dataset Creation 152

Table 4.5 Distribution of 153
Table 4.6 Summary of 162
Table 4.7 Summary of Audio Preprocessing Results 164
Table 4.8 Results of Balanced-Confident Dataset Split.................................... 164
Table 4.9 Format of Extracted MFCC Feature Data... 168
Table 4.10 Distribution of Labels in the MFCC Feature File 169
Table 4.11 Summary of 172
Table 4.12 Main Configuration of the 173
Table 4.13 Summary of Results from 10 Runs for Each Scenario 178
Table 4.14 Evaluation Results for Scenario S01 Over 10 Runs......................... 184
Table 4.15 Evaluation Results for Scenario S02 from 10 Runs 190
Table 4.16 Evaluation Results for Scenario S03 from 10 Runs 198
Table 4.17 Evaluation Results for Scenario S04 from 10 Runs 204
Table 4.18 Evaluation Results for Scenario S05 Over 10 Runs......................... 210
Table 4.19 Evaluation Results for Scenario S06 over 10 Runs.......................... 216
Table 4.20 Evaluation Results for Scenario S07 over 10 Runs.......................... 223
Table 4.21 Evaluation Results for Scenario S08 in 10 Runs.............................. 226
Table 4.22 Evaluation Results for Scenario S09 in 10 Runs.............................. 229
Table 4.23 Evaluation Results for Scenario S10 in 10 Runs.............................. 232
Table 4.24 Summary of Criteria for Selecting the Best Model Based on the Results of 10 Runs 235
Table 4.25 Interpretation of 240
Table 4.26 System Prototype Implementation Flowchart.................................. 243