UNDERGRADUATE THESIS

COMPARISON OF MANUAL AND SYNTHETIC DATA AUGMENTATION BASED ON VISION TRANSFORMER IN LUNG DISEASE CLASSIFICATION FROM RESPIRATORY SOUNDS

ADE RIZKY PANJAITAN
NPM 22081010091

THESIS ADVISORS
Fetty Tri Anggraeny, S.Kom., M. Kom
Eva Yulia Puspaningrum, 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

COMPARISON OF MANUAL AND SYNTHETIC DATA AUGMENTATION BASED ON VISION TRANSFORMER IN LUNG DISEASE CLASSIFICATION FROM RESPIRATORY SOUNDS

By:
ADE RIZKY PANJAITAN
NPM. 22081010091

Approved to proceed to the Thesis Examination

Approved by,

Coordinator of Informatics Study Program
Faculty of Computer Science

Dr. Intan Yuniar Purbasari, S.Kom.,M.Sc.
NIP. 19800602 2025212 029

APPROVAL SHEET

COMPARISON OF MANUAL AND SYNTHETIC DATA AUGMENTATION BASED ON VISION TRANSFORMER IN LUNG DISEASE CLASSIFICATION FROM RESPIRATORY SOUNDS

By:
ADE RIZKY PANJAITAN
NPM. 22081010091

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 July 10, 2026.

Approved,

Fetty Tri Anggraeny, S.Kom., M. Kom ..... **(Advisor I)**
NIP. 19820211 2021212 005

Eva Yulia Puspaningrum, S.Kom., M.Kom ..... **(Advisor II)**
NIP. 19890705 2021212 002

Dr. Faisal Muttaqin, S.Kom, M.T ..... **(Head Assessor)**
NIP. 19851231 2021211 009

M. Muharrom Al Haromainy, S.Kom., M.Kom ..... **(Assessor I)**
NIP. 19950601 202203 1 006

Acknowledge by,
Dean of the Faculty of Computer Science

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

STATEMENT OF ORIGINALITY

I am the undersigned:

Student Name	: Ade Rizky Panjaitan
NPM	: 22081010091
Degree Program	: Bachelor (S1)
Study Program	: Informatics
Faculty	: Faculty of Computer Science

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

And I declare that this scientific document is free from elements of plagiarism. If in the future indications of plagiarism are found in this Thesis, I am willing to accept sanctions in accordance with the applicable laws and regulations.

Thus, I made this statement without any coercion from anyone and to be used as it should.

Surabaya, July 23rd, 2026
Declarant,



ADE RIZKY PANJAITAN
NPM. 22081010091

ABSTRACT

Student Name / NPM:	Ade Rizky Panjaitan / 22081010091
Thesis Title:	COMPARISON OF MANUAL AND SYNTHETIC DATA AUGMENTATION BASED ON VISION TRANSFORMER IN LUNG DISEASE CLASSIFICATION FROM RESPIRATORY SOUNDS
Supervisor:	1. Fetty Tri Anggraeny, S.Kom. M.Kom
	2. Eva Yulia Puspaningrum, S.Kom., M.Kom

Class imbalance in respiratory sound datasets may cause classification models to favor majority classes and underperform on minority classes. This study investigates the effects of manual augmentation, MelGAN-based synthetic augmentation, and their combination on a Vision Transformer for classifying four respiratory sound classes in the ICBHI 2017 dataset: normal, crackle, wheeze, and both. Audio recordings were segmented, band-pass filtered, amplitude-normalized, resampled, standardized in duration, and converted into 224 × 224 log-Mel spectrograms. A ViT-B/16 model pretrained on ImageNet was trained through partial fine-tuning of the final six encoder blocks. Experiments covered original data, manual augmentation, synthetic MelGAN augmentation, and hybrid augmentation under patient-wise splits of 60:20:20, 70:15:15, and 80:10:10. The best model was selected using validation macro-F1. Synthetic MelGAN augmentation with a 70:15:15 split achieved the best performance, with 68.76% accuracy, 57.90% macro sensitivity, 85.50% macro specificity, and 59.99% macro-F1. Normal sounds were recognized most reliably, whereas crackle and both remained challenging. A supporting segment-level analysis yielded a macro-F1 of 58.74%. Overall, MelGAN augmentation provided the most balanced performance, while patient-wise splitting offered a stricter assessment of generalization to unseen patients.

Keywords **:** Data Augmentation, ICBHI 2017, MELGAN, Respiratory Sounds, Vision Transformer

ACKNOWLEDGEMENTS

Praise be to Allah SWT for all His graces, guidance, and gifts to the author so that the thesis proposal with the title **"Comparison Of Manual And Synthetic Data Augmentation Based On Vision Transformer In Lung Disease Classification From Respiratory Sounds"** can be completed properly.

The author would like to thank Mrs. Fetty Tri Anggraeny, S.Kom., M. Kom. as Supervisor I and Mrs. Eva Yulia Puspaningrum, S.Kom., M.Kom. as Supervisor II who are willing to take their time to provide guidance, advice and motivation to the author. In addition, during the preparation of the thesis proposal, the author also received a lot of assistance from various parties. For this the author would like to thank the:

1. Mrs. Prof. Dr. Ir. Novirina Hendrasarie, M.T. as the Dean of the Faculty of Computer Science, National Development University "Veteran" East Java.
2. Mrs. Dr. Intan Yuniar Purbasari, S.Kom., M.Sc. as the Coordinator of the Informatics Study Program, Faculty of Computer Science, National Development University "Veteran" East Java.
3. The author wishes to express sincere gratitude to Mr. Chrystia Aji Putra, S.Kom., M.Kom., the academic advisor, for the guidance, assistance, support, and attention provided throughout the academic journey, from the beginning of the program until the completion of studies.
4. Mr. Dr. Faisal Muttaqin, S.Kom., M.T., and Muhammad Muharrom Al Haromainy, S.Kom., M.Kom., served as examiners, providing evaluations, feedback, critiques, and suggestions throughout the thesis examination process. Their contributions and guidance have been invaluable to the author in refining the research and finalizing this thesis.
5. As a manifestation of profound love, devotion, and the highest gratitude, the author dedicates this to the beloved family, especially both parents, as well as the author's older sister and older brother. Endless thanks for all the financial support, moral encouragement, limitless patience, as well as unceasing prayers that always accompany every step even though the obstacles and the

journey of this thesis are not easy. Without their love, sacrifices, and support, the author would not be able to complete the compilation of this thesis well.

6. Special thanks and appreciation are dedicated to the author's closest friends from (BTS): Habib Nurrohmad Sugiharto, Ananda Putra WR, M. Fajar Saputra, Adhen Chandra, Rhimba Aulia, Cinta Ramayanti, and Dela Ayu. Thank you for being a sanctuary for sharing academic burdens and laughter amidst the demanding pace of university life. Your endless emotional support, patience, and the countless experiences we shared have been an extraordinary source of energy in completing this thesis. This friendship and its invaluable lessons remain one of the author's greatest college achievements. May God always bless your paths with success and happiness.
7. Distinct appreciation is also conveyed by the author to Rafie Najwan Anjas, who has helped a lot, supported, given advice as well as his effort and help indirectly in the working process up to the final revision stage of the thesis. As well as to Ama Maulidatul Khairah Thank you for all the help, support, and trust so that the author was able to hold on until the end and complete this thesis even though what was gone through was not easy and slightly tested patience. thank you for the endless support that is always given; even though sometimes also urging so that this thesis is immediately rushed, yet remaining consistent in giving support as well as spirit to chase every next checkpoint.
8. An honorable mention, is conveyed by the author to Azizah who has helped a lot, motivated, and supported from the beginning to the end of the working process. Even though often emotional and fed up yet still accompanying and giving spirit in the working process from the beginning to the end, even though this thesis process is far from the word smooth, testing patience, and clearly not oneshot. including lending an account for the necessity of executing the research topic trial, thank you for the patience, sincerity, up to her effort to always be bothered and for the time up to the energy spared.
9. A special shoutout is conveyed to the author's silat teammates for their direct and indirect help, moral support, and for being a safe space for thesis grievances. Thank you for the constant motivation and occasional roasting to push the author to immediately finish this S.Kom. degree. A specific

appreciation goes to Niam Joestar, who stood by the author from the very first page until its completion. Thank you all for the time and energy dedicated to seeing this thesis through.

10. The author conveys sincere gratitude to all members of HIMATIFA as an extraordinary platform to learn about responsibility, teamwork, leadership, and self-development outside the classroom. Deep appreciation is also directed to BLJ Naraya, for the endless support and togetherness that have become an inseparable part of the author's personal growth. Furthermore, thank you to Keluarga Mahasiswa (KM) Informatika 2022 for the support, spirit, and shared struggle passing through the ups and downs of the university period until achieving the S.Kom.
11. A well-deserved recognition is also conveyed by the author to the band Perunggu for their works that faithfully accompanied this thesis process. Through the songs '33x' and 'Amalan Baik', especially the lyric "melamban bukanlah hal yang tabu...", the author was made aware that no matter how small the progress, it remains an achievement. Their works embraced the author in times of exhaustion, muffled the feeling of being left behind by others, and became a reminder to keep fighting while surrendering to the Almighty until the author successfully completed what had been started.
12. Finally, the author deeply thanks everyone who cannot be mentioned individually, yet whose direct or indirect support, prayers, and assistance made this thesis possible. May your kindness be rewarded with the best of happiness and success.

Realizing the limitations and shortcomings of this thesis, the author openly welcomes constructive criticism for future improvements. Ultimately, the author hopes this work provides valuable insights for readers and serves as a meaningful learning experience personally.

Surabaya, July 23rd, 2026

Author

Table of Contents

ABSTRACT ..... **iv**
ACKNOWLEDGEMENTS ..... **v**
Table of Contents ..... **viii**
List of Figures ..... **xii**
List of Tables ..... **xv**
CHAPTER I INTRODUCTION **1**
1.1. Background 1
1.2. Problem Formulation 4
1.3. Research Objectives 5
1.4. Research Benefits 5
1.5. Scope and Limitations 6
CHAPTER II LITERATURE REVIEW **7**
2.1. Prior Research 7
2.2. The Concepts of AI, Machine Learning, and Deep Learning 9
2.3. The Human Respiratory System and Respiratory Sounds 12
2.4. Acoustic Features of Respiratory Sounds 15
2.4.1. Time Domain Features 16
2.4.2. Frequency Domain Features 16
2.4.3. Cepstral Features 17
2.5. Audio to Image Representation 18
2.5.1. Spectrogram 18
2.5.2. Mel-Spectrograms 20
2.5.3. Log-Mel Spectrogram 22
2.5.4. Normalisasi dan Pembentukan Citra *log-Mel spectrogram* 23
2.6. Preprocessing of Respiratory Sounds 24
2.6.1. Band pass Filter 24
2.6.2. Amplitude Normalization 26
2.6.3. Audio Resampling 28
2.7. Data Augmentation 30
2.7.1. Manual Augmentation (Audio Level) 30
2.7.2. Synthetic Augmentation (Audio Level / MelGANs) 32

2.8. The Vision Transformer Methodology (ViT) 34
2.8.1. The Conceptual Framework of Self Attention and Patch Embedding .36
2.8.2. Architecture and Components of the Vision Transformer 39
2.9. Evaluation Metrics 42
2.9.1. Confusion Matrix 43
2.9.2. Accuracy 43
2.9.3. Recall (Sensitivity) 43
2.9.4. Specificity 44
2.9.5. F1-Score 44
CHAPTER III SYSTEM DESIGN AND IMPLEMENTATION 45
3.1. Research Methodology 45
3.2. Data Collection 46
3.3. Audio Segmentation and Labeling 49
3.4. Dataset Partitioning 52
3.5. Data Preprocessing 55
3.5.1. Band Pass Filter 100 to 2000 Hz 55
3.5.2. Amplitude Normalization 56
3.5.3. Resampling to 4000 Hz 58
3.6. Audio Data Augmentation 59
3.6.1. Manual Augmentation 61
3.6.2. Synthetic Augmentation Utilizing MelGAN 64
3.6.3. Integration of Manual and Synthetic Augmentation 67
3.7. Log-Mel Spectrogram Extraction 72
3.7.1. Short-Time Fourier Transform (STFT) Configuration 73
3.7.2. Mel Spectrogram Formation 73
3.7.3. Conversion to Log-Mel Spectrogram 74
3.8. Normalization and Image Construction 76
3.8.1. Log-Mel Spectrogram Intensity Normalization 76
3.8.2. Vision Transformer Input Image Formation 77
3.9. Vision Transformer (ViT) Model Design 78
3.9.1. Vision Transformer (ViT) Model Architecture 79
3.9.2. Training Configuration 82

3.9.3. Hyperparameter Optimization 85
3.9.4. Training and Inference Procedure 88
3.10. Testing Scenarios 89
3.10.1. Testing Scenario Design 90
3.10.2. Augmentation Testing Scenarios 91
3.10.3. Testing Execution and Evaluation Protocol 95
3.11. Model Evaluation 96
3.11.1. Confusion Matrix Construction 97
3.11.2. Classification Metric Calculation 97
3.11.3. Comparative Analysis Among Scenario Combinations 98
3.12. Research Tools and Environment 99
3.12.1. Hardware 99
3.12.2. Software 100
CHAPTER IV RESULTS AND DISCUSSION 101
4.1. Testing Implementation 101
4.2. Dataset Distribution and Augmentation Results 105
4.2.1. 60:20:20 Split 110
4.2.2. 70:15:15 Split 111
4.2.3. 80:10:10 Split 112
4.2.4. Training Set Distribution After Augmentation 113
4.3. Visualization of Augmentation Results 114
4.4. Hyperparameter Optimization and Model Training Results 118
4.4.1. 60:20:20 Split 118
4.4.2. 70:15:15 Split 120
4.4.3. 80:10:10 Split 121
4.4.4. Training Curve Analysis 123
4.5. Model Testing Results on the Test Set 126
4.5.1. 60:20:20 Split 126
4.5.2. 70:15:15 Split 128
4.5.3. 80:10:10 Split 130
4.6. Comparison of All Experimental Combinations 132
4.6.1. Comparison of Each Augmentation Strategy 134

4.6.2. Comparison for Each Ratio 136
4.6.3. Determination of the Best Combination 138
4.7. Misclassification Analysis 139
4.8. Discussion of Results in Relation to the Research Objectives 140
4.9. Supplementary Analysis of the Segment-Level Split 142
4.9.1. Data Distribution in the Segment-Level Split 143
4.9.2. Configuration of the Segment-Level Split Analysis 147
4.9.3. Model Training Results in the Segment-Level Split 148
4.9.4. Segment-Level Split Testing Results on the Test Set 150
4.9.5. Comparison of Patient-Wise Split and Segment-Level Split Results 153
4.9.6. Summary and Interpretation of the Segment-Level Split Results 155
4.10. Comparison with Previous Research 157
4.11. Research Limitations 161
CHAPTER V CONCLUSION AND RECOMMENDATIONS 163
5.1. Conclusion 163
5.2. Recommendations 164
BIBLIOGRAPHY 166
APPENDIX 171

List of Figures

Figure 2.1 The Hierarchy of Artificial Intelligence, Machine Learning, and Deep Learning. [30] 10
Figure 2.2 The taxonomy of supervised and unsupervised machine learning algorithms. [32] 11
Figure 2.3 General Deep Learning Architecture, and several primary architectures within deep learning: CNN, RNN, LSTM, Autoencoder, GAN, and Transformer Neural Network. [33] 12
Figure 2.4 The Human Respiratory System. [37] 13
Figure 2.5 A comparison between normal and abnormal airways. [36] 14
Figure 2.6 Examples waveforms of respiratory sounds from the ICBHI 2017 dataset 14
Figure 2.7 Waveforms of crackles and wheezes within the time domain 16
Figure 2.8 An example of a linear spectrogram derived from the ICBHI 2017 dataset 19
Figure 2.9 An example of a log Mel spectrogram derived from ICBHI 2017 dataset 21
Figure 2.10 Conceptual operational schematic detailing the Vision Transformer framework. [35] 35
Figure 3.1 Research Methodology Flowchart 45
Figure 3.2 Example annotation and audio data within the ICBHI 2017 dataset 48
Figure 3.3 Exemplary log Mel spectrogram illustrating respiratory sounds categorized under the Crackle and Wheeze classes 49
Figure 3.4 Flowchart detailing the segmentation and labeling procedures for respiratory sounds within the ICBHI 2017 dataset 50
Figure 3.5 Exemplary audio following segmentation and with zero padding application 51
Figure 3.6 Flowchart delineating the dataset partitioning based upon the patient wise split approach alongside the augmentation application upon the training set preceding the log Mel spectrogram image extraction 54
Figure 3.7 Flowchart detailing data preprocessing 55

Figure 3.8 Waveforms preceding and succeeding the application of 100 to 2000 Hz band pass filter 56
Figure 3.9 Waveforms preceding and succeeding Root Mean Square based amplitude normalization 57
Figure 3.10 Density differences preceding and succeeding the 4000 Hz resampling following a 50 ms magnification 58
Figure 3.11 Flowchart detailing the manual alongside the synthetic audio data augmentation within the research study 60
Figure 3.12 Flowchart of MelGAN model training on respiratory sound data 65
Figure 3.13 Flowchart illustrating the MelGAN-based synthetic augmentation data generation process 66
Figure 3.14 Log-Mel spectrogram of the original data (a) and the pitch shifting augmentation result (b) 71
Figure 3.15 Log-Mel spectrogram of the time stretching result (c) and the MelGAN synthetic result (d) 71
Figure 3.16 Flowchart depicting the log-Mel spectrogram extraction process from 5-second audio segment data 75
Figure 3.17 Workflow diagram of the Vision Transformer for log-Mel spectrogram image classification 79
Figure 3.18 Flowchart of Testing Scenarios 95
Figure 4.1 Training Data Distribution per Class Across Three Patient-Wise Split Ratios and Four Augmentation Scenarios 109
Figure 4.2 Training Set Distribution per Class for the 60:20:20 Patient-Wise Split Ratio Across Four Augmentation Scenarios 110
Figure 4.3 Training Set Distribution per Class for the 70:15:15 Patient-Wise Split Ratio Across Four Augmentation Scenarios 111
Figure 4.4 Training Set Distribution per Class for the 70:15:15 Patient-Wise Split Ratio Across Four Augmentation Scenarios 112
Figure 4.5 Comparison of waveform and log-Mel spectrogram images between original data and manual augmentation results 115
Figure 4.6 Comparison of waveforms and log-Mel spectrogram images between original data and generative synthetic augmentation results MelGAN Model 116

Figure 4.7 Comparison of training and validation macro-F1 60:20:20 split 119
Figure 4.8 Comparison of training and validation macro-F1 70:15:15 split 121
Figure 4.9 Comparison of training and validation macro-F1 80:10:10 split 122
Figure 4.10 Training and validation loss curves for the 60:20:20 split: (a) original data, (b) manual augmentation, (c) MelGAN, and (d) hybrid 123
Figure 4.11 Training and validation loss curves for the 70:15:15 split: (a) original data, (b) manual augmentation, (c) MelGAN, and (d) hybrid 124
Figure 4.12 Training and validation loss curves for the 80:10:10 split: (a) original data, (b) manual augmentation, (c) MelGAN, and (d) hybrid 125
Figure 4.13 Comparison of accuracy, macro sensitivity, macro specificity, and Macro-F1 for the 60:20:20 split 128
Figure 4.14 Comparison of accuracy, macro sensitivity, macro specificity, and Macro-F1 for the 70:15:15 split 130
Figure 4.15 Comparison of accuracy, macro sensitivity, macro specificity, and Macro-F1 for the 80:10:10 split 131
Figure 4.16 Comparison of Average Performance for Each Strategy Across the Three Patient-Wise Split Ratios 135
Figure 4.17 Comparison of Macro-F1 for Each Strategy Across the Three Patient-Wise Split Ratios 137
Figure 4.18 Confusion matrix of the MelGAN combination 70:15:15 split 139
Figure 4.19 Distribution of Class Proportions 70:15:15 Segment-Level Split ... 144
Figure 4.20 Comparison of the Training Set Distribution Before and After MelGAN Synthetic Augmentation 146
Figure 4.21 Model Training Curve for the 70:15:15 Segment-Level Split 149
Figure 4.22 Confusion Matrix of the Segment-Level Split Testing Results 151
Figure 4.23 Comparison of the Evaluation Metrics Between the Patient-Wise Split and Segment-Level Split 155

List of Tables

Table 2.1 Representative Confusion Matrix Configuration 43
Table 3.1 Audio File and Annotation Format within the ICBHI 2017 Dataset 47
Table 3.2 Attributes derived from an exemplary text annotation document 48
Table 3.3 Exemplary label determination derived for segmentation outcomes 51
Table 3.4 Dataset Partitioning Scenarios 53
Table 3.5 Data distribution respiratory sound segment across every split 53
Table 3.6 Distribution augmented respiratory sound data allocation for each scenario 69
Table 3.7 Training Configuration 84
Table 3.8 Summary of Testing Scenarios 92
Table 3.9 Research Hardware Specifications 100
Table 4.1 Testing Implementation Workflow 103
Table 4.2 Summary of Testing Scenarios 104
Table 4.3 Patient and Segment Distribution for Each Patient-Wise Split 105
Table 4.4 Final Training Set Distribution for Each Experiment Combination ... 107
Table 4.5 Model Optimization and Validation Results the 60:20:20 Split 118
Table 4.6 Model Optimization and Validation Results the 70:15:15 Split 120
Table 4.7 Model Optimization and Validation Results the 80:10:10 Split 121
Table 4.8 Model Testing Results for the 60:20:20 Split 127
Table 4.9 Model Testing Results for the 70:15:15 Split 129
Table 4.10 Model Testing Results for the 80:10:10 Split 131
Table 4.11 Comparison of All Experimental Combinations 133
Table 4.12 Average Performance of Each Strategy Across the Three Patient-Wise Split Ratios 135
Table 4.13 Best Strategy for Each Ratio Based on Macro-F1 137
Table 4.14 Ranking of the Best Combinations Based on Macro-F1 138
Table 4.15 Original Data Distribution in the 70:15:15 Segment-Level Split 143
Table 4.16 Audit Results of Overlap in the Segment-Level Split 145
Table 4.17 Training Set Distribution After MelGAN Synthetic Augmentation . 145
Table 4.18 Summary of the Segment-Level Split Analysis Configuration 147
Table 4.19 Model Training Results in the Segment-Level Split 149

Table 4.20 Evaluation Metrics per Class in the Segment-Level Split 151
Table 4.21 Summary of the Segment-Level Split Testing Results 152
Table 4.22 Comparison of Patient-Wise and Segment-Level Split Results 153
Table 4.23 Comparison of Sensitivity and F1-Score Between the Patient-Wise and Segment-Level Split Results 154
Table 4.24 Comparison of Research Results with Previous Studies 158