



**CONVOLUTIONAL FEATURE OPTIMIZATION FOR
LUNG DISEASE CLASSIFICATION ON
IMAGE DATA**

THESIS REPORT

DAVID GINOLA GUTERRES

41522010208

**UNIVERSITAS
MERCU BUANA**

**INFORMATICS ENGINEERING STUDY PROGRAM
FACULTY OF COMPUTER SCIENCE
MERCU BUANA UNIVERSITY
JAKARTA
2025**



**CONVOLUTIONAL FEATURE OPTIMIZATION FOR
LUNG DISEASE CLASSIFICATION ON
IMAGE DATA**

THESIS REPORT

DAVID GINOLA GUTERRES

41522010208

**UNIVERSITAS
MERCU BUANA**

Submitted as one of the requirements to obtain a bachelor's degree

**INFORMATICS ENGINEERING STUDY PROGRAM
FACULTY OF COMPUTER SCIENCE
MERCU BUANA UNIVERSITY
JAKARTA
2025**

OWN WORK STATEMENT PAGE

I, the undersigned below :
Name : David Ginola Guterres
Student Number : 41522010208
Faculty/Study Program : Computer Science/Informatics Engineering
Thesis Title : Convolutional Feature Optimization for Lung
Disease Classification on Image Data

I declare that this Research Proposal is my own work and not plagiarized, and all sources, both cited and referred to, have been accurately acknowledged. If it is found that there is plagiarism in my Research Proposal, I am ready to face the academic sanctions applicable at Mercu Buana University.



UNIVERSITAS
MERCU BUANA

Jakarta, 12th January 2026



David Ginola Guterres

TURNITIN TEST RESULT CERTIFICATE PAGE



072.423.4.03.01

PERNYATAAN SIMILARITY CHECK /SIMILARITY CHECK STATEMENT

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh

The undersigned, hereby declare that the scientific work written by

Nama /Name	: David Ginola Guterres
NIM /Student ID Number	: 41522010208
Program Studi/Study Program	: Teknik Informatika

Dengan Judul Tugas Akhir

The title

"Convolutional Feature Optimization for Lung Disease Classification on Image Data"

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

Similarity check have been carried out with the Turnitin system on the date

9 Januari 2026

dengan nilai persentase sebesar :

with a percentage value of

11%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di Fakultas Ilmu Komputer

Universitas Mercu Buana. *declared to meet standards to accordance with applicable regulations at the Faculty of Computer Science, Universitas Mercu Buana.*

File hasil cek similarity turnitin:

Turnitin similarity report file

https://drive.google.com/file/d/1FYg1sT6nKtGBd5ehv4YqT4y2Z2W3/view?usp=drive_link



Jakarta, 9 Januari 2026
Admin Turnitin Fasilkom UMB

Agung Prawoto

Agung Prawoto, S.Kom., B.Sc

NIK : 322970503

Fakultas Ilmu Komputer

KAMPUS MENARA BHAKTI

Jl. Raya Meruya Selatan No. 1 Kembangan, Jakarta Barat 11650

Telp. 021-5840816 (Hunting), Paw : 5700 Fax. 021-5840813

<http://www.mercubuana.ac.id>, e-mail : fasilkom@mercubuana.ac.id

APPROVAL PAGE

This Thesis Report is submitted by :

Name : David Ginola Guterres
Student Number : 41522010208
Faculty/Study Program : Computer Science/Informatics Engineering
Thesis Title : Convolutional Feature Optimization for Lung
Disease Classification on Image Data

It was successfully defended at the hearing on January 19th, 2025, and accepted as part of the requirements necessary to obtain a Bachelor's degree in the Informatics Engineering Study Program, Faculty of Computer Science, Mercu Buana University.

Approved by:
Supervisor



Dr. Hadi Santoso, S.Kom., M.Kom.

NIDN/NUPTK: 0225067701

UNIVERSITAS

Jakarta, 26th January 2026

Knowing,

MERCU BUANA

Dean of Faculty of Computer Science

Head of Informatics Department



Dr. Bambang Jokonowo, S.Si., MTI
NIDN/NUPTK: 0320037002



Dr. Hadi Santoso, S.Kom., M.Kom.
NIDN/NUPTK: 0225067701

PREFACE

All praise and gratitude are due to Almighty God Jesus Christ, whose grace and guidance have enabled the author to complete this thesis proposal entitled **“Convolutional Feature Optimization for Lung Disease Classification on Image Data.”** This proposal is prepared as one of the academic requirements for obtaining a bachelor’s degree in the S-1 International Class Study Program, Department of Informatics, Faculty of Computer Science, Universitas Mercu Buana Jakarta.

The author fully realizes that, as a human being with limited knowledge and experience, this thesis proposal is not free from weaknesses and limitations. Throughout the preparation process, various challenges were encountered, yet they became valuable learning experiences that contributed to both academic and personal growth. The completion of this proposal would not have been possible without the guidance, encouragement, and support received from many parties during the research and writing process. Special gratitude is extended to **Dr. Hadi Santoso, S.Kom., M.Kom.**, my supervisor, for providing invaluable guidance, advice, and encouragement throughout this journey. On this occasion, I express gratitude to all those who contributed to the successful completion of this research, with special mention to:

1. Prof. Dr. Andi Adriansyah, M. Eng as the Rector of Mercu Buana University,
2. Dr. Bambang Jokonowo, S.Si., MTI as the Dean of the Faculty of Computer Science,
3. Dr. Hadi Santoso, S.Kom., M.Kom as the Head of the Department of Informatics Engineering,
4. Dr. Hadi Santoso, S.Kom., M.Kom., as the Thesis Advisor who has dedicated time, effort, and priceless counsel has been a cornerstone for my journey in completing this thesis and also as the lecturer in charge of the Research Methodology Major.
5. Afiyati, Dr., S.Si, MT as Deputy Dean of the International Department of Informatics Engineering, Universitas Mercu Buana.
6. Ilham Nugraha S.Kom, M.Cs, as Secretary of International Department of Informatics Engineering, Universitas Mercu Buana.
7. I am also deeply thankful to my parents who have provided support and consistently offered prayers to the author.
8. I would like to acknowledge and thank all my friends especially my four classmate for supporting me either mentally or physically.
9. I would also like to acknowledge and thank the various parties and individuals who have given motivation and beneficial knowledge to the author.

The author humbly recognizes the imperfections in this thesis proposal, acknowledging the limitations in knowledge and experience. The author openly welcomes suggestions, feedback, and constructive criticism for the improvement of this thesis proposal. It is hoped that this work can contribute positively to the development of knowledge, particularly in the field of image classification and deep learning applications. The author sincerely apologizes for any shortcomings contained in this proposal and hopes it may be of benefit to readers and future researchers.

Jakarta, 12th January 2026



David Ginola Guterres



**PUBLICATION CONSENT STATEMENT PAGE FOR FINAL PROJECT
IN THE UMB REPOSITORY**

As a member of the academic community at Universitas Mercu Buana, I the undersigned below :

Name : David Ginola Guterres
Student Number : 41522010208
Faculty/Study Program : Computer Science/Informatics Engineering
Thesis Title : Convolutional Feature Optimization for Lung Disease Classification on Image Data

For the advancement of knowledge, hereby grant permission and approve the provision of the Non-Exclusive Royalty-Free Right to Mercu Buana University for my scholarly work titled above, along with its accompanying materials (if necessary).

With this Non-Exclusive Royalty-Free Right, Mercu Buana University is authorized to store, convert/format, manage in the form of a database, maintain and publish my Internship Report/Thesis/Dissertation, while ensuring the inclusion of my name as the author/creator and copyright owner.

I make this statement sincerely for the record.

UNIVERSITAS
MERCU BUANA

Jakarta, 12th January 2026



David Ginola Guterres

ABSTRAK

Nama : David Ginola Guterres
NIM : 41522010208
Program Studi : Informatics Engineering
Judul Tugas Akhir : Convolutional Feature Optimization For Lung
Disease Classification on Image Data
Pembimbing : Dr. Hadi Santoso, S.Kom., M.Kom.

Penyakit paru-paru seperti pneumonia, tuberkulosis, dan penyakit paru obstruktif kronis (PPOK) merupakan tantangan kesehatan global yang signifikan, yang memerlukan metode diagnostik yang akurat dan efisien. Jaringan Saraf Konvolusional (CNN) telah menunjukkan kemampuan luar biasa dalam analisis gambar medis, terutama dalam klasifikasi penyakit menggunakan sinar-X dada dan pemindaian tomografi komputer (CT). Studi ini mengeksplorasi optimasi model CNN untuk klasifikasi penyakit paru-paru dengan menerapkan teknik prapemrosesan canggih, penyesuaian hiperparameter, dan augmentasi dataset. Menggunakan dataset yang tersedia secara publik, berbagai arsitektur CNN, termasuk ResNet50, Efficient Net, dan Retina Net, dievaluasi untuk menentukan efektivitasnya dalam membedakan antara kategori penyakit paru-paru yang berbeda. Studi ini mencapai akurasi optimal hingga 98,89%, menunjukkan potensi pembelajaran mendalam dalam membantu tenaga medis dalam deteksi dini dan akurat penyakit paru-paru. Arah penelitian masa depan meliputi integrasi model hibrida, seperti CNN-LSTM dan teknik pembelajaran ensambel, untuk lebih meningkatkan akurasi dan ketahanan klasifikasi.

Kata Kunci : (CNN), Sinar-X, Pemindaian Tomografi Terkomputasi (CT), ResNet50, EfficientNet, RetinaNet, CNN-LSTM.

ABSTRACT

Name : David Ginola Guterres
NIM : 41522010208
Study program : Informatics Engineering
Research Proposal Title : Convolutional Feature Optimization For Lung Disease Classification on Image Data
Supervisor : Dr. Hadi Santoso, S.Kom., M.Kom.

Lung diseases such as pneumonia, tuberculosis, and chronic obstructive pulmonary disease (COPD) present significant global health challenges, requiring accurate and efficient diagnostic methods. Convolutional Neural Networks (CNNs) have shown remarkable capabilities in medical image analysis, particularly in disease classification using chest X-rays and computed tomography (CT) scans. This study explores the optimization of CNN models for lung disease classification by implementing advanced preprocessing techniques, hyperparameter tuning, and dataset augmentation. Using publicly available datasets, various CNN architectures, including ResNet50, Efficient Net and Retina Net are evaluate to determine their effectiveness in distinguishing between different lung disease categories. The study achieves an optimized accuracy of up to 98.89%, demonstrating the potential of deep learning in assisting medical professionals with early and precise lung disease detection. Future research directions include integrating hybrid models, such as CNN-LSTM and ensemble learning techniques, to further improve classification accuracy and robustness.

Keywords : (CNNs), X-rays, Computed Tomography (CT) scans, ResNet50, EfficientNet, RetinaNet, CNN-LSTM.

TABLE OF CONTENTS

COVER PAGE	0
TITLE PAGE	i
OWN WORK STATEMENT PAGE	ii
TURNITIN TEST RESULT CERTIFICATE PAGE	iii
APPROVAL PAGE	iv
PREFACE.....	v
PUBLICATION CONSENT STATEMENT PAGE FOR FINAL PROJECT IN THE UMB REPOSITORY	vii
ABSTRAK	viii
ABSTRACT	ix
TABLE OF CONTENTS	x
LIST OF TABLES	xii
LIST OF FIGURES	xiii
APPENDIX LIST.....	xiv
CHAPTER I INTRODUCTION	1
1.1 Background	1
1.2 Problem Statement.....	4
1.3 Research Objective	4
1.4 Research Benefits.....	4
1.5 Research Scope	5
CHAPTER II LITERATURE REVIEW	6
2.1 Main Theory.....	6
2.1.1 Lung disease classification	6
2.1.2 Convolutional Neural Networks (CNNs) in Medical Imaging	7
2.1.3 Optimization Techniques in CNNs.....	8
2.2 Previous Research	9
CHAPTER III METHODOLOGY	21
3.1 Research Framework	21
3.2 Data Collection	23
3.3 Data Pre-processing	24
3.4 CNN Model Design and Training.....	26
3.5 Model Evaluation.....	28

CHAPTER IV RESULT AND DISCUSSION	29
4.1 Dataset.....	29
4.2 Preprocessing	30
4.3 Model Architecture	32
4.4 Model Performance.....	35
4.5 Testing.....	38
4.6 Result Visualization	40
CHAPTER V CONCLUSION AND SUGGESTION	41
5.1 Conclusion	41
5.2 Suggestion and Recommendation.....	42
REFERENCES.....	43
ATTACHMENT	46



LIST OF TABLES

Table 2. 1 Previous Research.....	9
Table 4. 1 Train Classification Report.....	37
Table 4. 2 Test Classification Report.....	40



LIST OF FIGURES

Figure 3. 1 Research Method Flowchart.....	23
Figure 4. 1 Normal Lung Images	29
Figure 4. 2 Lung Opacity Images	30
Figure 4. 3 Pneumonia Virus Images.....	30
Figure 4. 4 User Parameters	31
Figure 4. 5 Data Generators Code.....	32
Figure 4. 6 Rest Net Classifier Code	33
Figure 4. 7 Retina Net FPN Classifier Code.....	34
Figure 4. 8 PCA-Rest Net Code.....	35
Figure 4. 9 RestNet Accuracy and Loss of Balance Dataset	35
Figure 4. 10 RetinaNet Accuracy and Loss of Balance Dataset	36
Figure 4. 11 RestNet Accuracy and Loss of Imbalance Dataset.....	36
Figure 4. 12 RetinaNet Accuracy and Loss of Imbalance Dataset	36
Figure 4. 13 Train Confusion Matrix.....	37
Figure 4. 14 Test Accuracy	38
Figure 4. 15 Test Confusion Matrix.....	39
Figure 4. 16 Result Visualization.....	40



APPENDIX LIST

Attachment 1. Assistance Card	46
Attachment 2. Page of Final Project Outcome Statement.....	47
Attachment 3. Journal Article Manuscript	48
Attachment 4. Curriculum Vitae	49
Attachment 5. IBM Certificate.....	50
Attachment 6. Statement and Transfer of Intellectual Property Rights	51

