UNIVERSITAS
GADJAH MADA
**An Empirical Analysis on Feature Complementarity and Fusion Depth in Multimodal Deep Learning for
Music Genre Classification**
Daniel Winston Mandela Tulung, Rifki Afina Putri, S.T., M.S., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

TABLE OF CONTENTS

TABLE OF CONTENTS **4**
LIST OF TABLES **6**
LIST OF FIGURES **7**
ABSTRACT **9**
CHAPTER 1
INTRODUCTION **1**
1.1 Research Background 1
1.2 Problem Formulation 3
1.3 Research Scope 4
1.4 Research Objective 4
1.5 Research Significance 4
CHAPTER 2
LITERATURE REVIEW **6**
CHAPTER 3
RESEARCH METHODOLOGY **11**
3.1 General Description of the Research 11
3.2 Data 14
3.2.1 Dataset Description 15
3.2.2 Track-Level Splitting and Segmentation Strategy 15
3.2.3 Data Normalization 16
3.3 Method and Model Design 17
3.3.1 Spatial Feature Extraction: Mel-Spectrograms 17
3.3.2 Statistical Audio Feature Extraction 18
3.3.3 Model Architecture and Fusion Implementation 22
3.3.3.1 Deep Learning Base Architectures 23
3.3.3.2 Fusion Strategies (The Ablation Targets) 24
3.4 Implementation and Ablation Strategy 27
CHAPTER 4 ..... **30**
IMPLEMENTATION **30**
4.1 Implementation Steps 30
4.2 Tools and Materials 33
4.3 Program Code and Explanation 34
4.3.1 Global Configuration and Hyperparameters 34

UNIVERSITAS
GADJAH MADA
**An Empirical Analysis on Feature Complementarity and Fusion Depth in Multimodal Deep Learning for
Music Genre Classification**
Daniel Winston Mandela Tulung, Rifki Afina Putri, S.T., M.S., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

4.3.2 CNN and MLP Feature-Extractor Branches..... 36
4.3.3 Fusion Wrapper Architectures (Early / Intermediate / Late).....38
4.3.4 Training Loop, Checkpointing, and Early Stopping..... 40
4.3.5 Majority-Vote Track-Level Evaluation..... 43
4.3.6 Programmatic Ablation-Matrix Construction..... 44
4.3.7 Repeated-Partition Robustness Extension.....45
4.3.8 Hyperparameter Grid Search for the Selected Configuration..... 46
4.4 Implementation Testing..... 49
4.4.1 Pipeline Validation (Smoke Testing).....49
4.4.2 Full-Scale Execution and Early-Stopping Behavior..... 53
4.4.3 Error Tracking and Resolution.....54
4.5 Code Documentation..... 56
CHAPTER 5..... 58
RESULTS AND DISCUSSION.....58
5.1 Overview of the Experimental Runs and Data Validity.....58
5.2 Aggregated Ablation Matrix Results..... 59
5.3 Effect of Fusion Depth.....65
5.4 Effect of Statistical Descriptor Subset Selection..... 67
5.5 Interaction Between Fusion Depth and Feature Subset.....69
5.6 Cross-Split Robustness of the Leading Configurations.....70
5.7 Hyperparameter Grid Search for the Selected Configuration.....73
5.8 The Effect of Track-Level Majority Voting.....75
5.9 Effect of Segment Length on Track-Level Performance.....76
5.10 Training Cost versus Performance of the Best Configurations.....80
5.11 Discussion: Addressing the Research Objectives.....83
CHAPTER 6..... 86
CONCLUSION.....86
6.1 Conclusion..... 86
6.2 Contributions..... 88
6.3 Limitations.....88
6.4 Recommendations for Future Work.....89
LIST OF ABBREVIATIONS AND GLOSSARY..... 90
REFERENCES.....94