



## **UNDERGRADUATE THESIS**

# **CLASSIFICATION OF JAVANESE DIALECTS USING A HYBRID CNN-TCN DEEP LEARNING ARCHITECTURE**

**DANIKA NAJWA ARDELIA**  
NPM 22081010103

### **THESIS ADVISORS**

Dr. Faisal Muttaqin, S.Kom, MT  
M. Muharrom Al Haromainy, S.Kom., M.Kom.

**MINISTRY OF HIGHER EDUCATION, SCIENCE, AND TECHNOLOGY  
VETERAN NATIONAL DEVELOPMENT UNIVERSITY OF EAST JAVA  
FACULTY OF COMPUTER SCIENCE  
INFORMATICS STUDY PROGRAM  
SURABAYA  
2026**

## APPROVAL SHEET

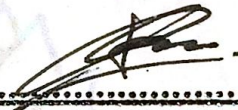
### CLASSIFICATION OF JAVANESE DIALECTS USING A HYBRID CNN-TCN DEEP LEARNING ARCHITECTURE

By:  
DANIKA NAJWA ARDELIA  
NPM. 22081010103

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 East Java National Veterans, on May 19, 2026:

Approved,

Dr. Faisal Muttagin, S.Kom, MT  
NIP. 19851231 202121 1 009



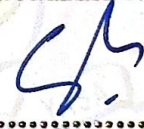
(Advisor I)

M. Muharrom Al Haromainy, S.Kom.,  
M.Kom  
NIP. 19950601 202203 1 006



(Advisor II)

Eva Yulia Puspaningrum, S.Kom,  
M.Kom  
NIP. 19890705 202121 2 002



(Head Assessor)

Yisti Vita Via, S.ST, M.Kom.  
NIP. 19860425 202121 2 001



(Assessor I)

Acknowledged by,

Dean of the Faculty of Computer Science



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

## **APPROVAL SHEET**

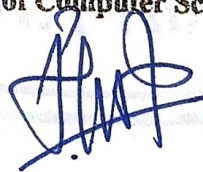
### **CLASSIFICATION OF JAVANESE DIALECTS USING A HYBRID CNN-TCN DEEP LEARNING ARCHITECTURE**

By:  
**DANIKA NAJWA ARDELIA**  
NPM. 22081010103

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.19800602 202521 2 029**

## STATEMENT OF ORIGINALITY

I am the undersigned:

Student Name : Danika Najwa Ardelia  
NPM : 22081010103  
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 educational 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, June 15, 2026  
Declarant,



**DANIKA NAJWA ARDELIA**  
NPM. 22081010103

## ABSTRACT

Student Name / NPM : Danika Najwa Ardelia/22081010103  
Thesis Title : Classification of Javanese Dialects Using A Hybrid  
CNN-TCN Deep Learning Architecture  
Supervisor : 1. Dr. Faisal Muttaqin, S.Kom, MT  
2. M. Muharrom Al Haromainy, S.Kom., M.Kom.

The decline in the use of regional languages in Indonesia, including Javanese, threatens the preservation of local cultural identity. Classifying Javanese dialects is challenging due to subtle variations in pitch and tempo. This study proposes a hybrid Convolutional Neural Network (CNN) and Temporal Convolutional Network (TCN) architecture to identify five Javanese dialects: Arekan, Pantura (Banten), Banyumas, Cirebon, and Solo-Yogya. The CNN front-end extracts spatial features from Log-Mel Spectrograms, while the TCN back-end models long-term temporal dependencies of the speech signals.

Data was collected from The Language Archive (TLA) and YouTube, then preprocessed through 16 kHz resampling, Voice Activity Detection (VAD), 10-second segmentation with 30% overlap, and feature extraction. Noise Injection and SpecAugment techniques were applied to enrich the data and prevent overfitting. The system was evaluated using Stratified Group 5-Fold Cross Validation, employing a Soft Voting probability aggregation mechanism to combine segment-level predictions into a final audio-level decision.

The results demonstrate that the CNN-TCN integration achieved an accuracy of 90.68% and an F1-Score of 0.9032. This hybrid architecture significantly outperformed the single CNN model (83.53%) and TCN model (86.37%). Furthermore, it effectively resolved cross-dialect classification confusion, notably improving the recognition of the Arekan dialect. Finally, the optimal classification model was successfully implemented into an interactive web-based application interface to facilitate dialect identification for end users.

**Keywords:** Javanese Dialect, Dialect Identification, Convolutional Neural Network, Temporal Convolutional Network, Log-Mel Spectrogram.

## ACKNOWLEDGEMENTS

All praise and gratitude are devoted to Allah SWT for His endless mercy, blessings, and guidance, which have granted the author the strength and opportunity to complete this undergraduate thesis entitled **“Classification of Javanese Dialects Using A Hybrid CNN-TCN Deep Learning Architecture.”**

This thesis was prepared as part of the requirements for obtaining a Bachelor of Computer Science degree in the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur.

The author acknowledges that the completion of this research would not have been possible without the invaluable support, motivation, prayers, and contributions of many individuals. Their assistance, whether in the form of academic guidance, moral encouragement, spiritual support, or practical help, played an essential role throughout the research process. Therefore, the author would like to sincerely express heartfelt appreciation and gratitude to:

1. Prof. Dr. Ir. Novirina Hendrasarie, M.T., Dean of the Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur.
2. Dr. Intan Yuniar Purbasari, S.Kom., M.Sc., Head of the Informatics Study Program.
3. Dr. Faisal Muttaqin, S.Kom., M.T., and M. Muharrom Al Haromainy, S.Kom., M.Kom., as thesis supervisors, for generously sharing their expertise, providing valuable insights, and offering continuous guidance that greatly contributed to the completion of this research.
4. Eva Yulia Puspaningrum, S.Kom., M.Kom., and Yisti Vita Via, S.ST., M.Kom., as thesis assessors, for their thoughtful evaluations, constructive feedback, and recommendations that helped refine and strengthen the quality of this thesis.
5. All lecturers of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, who have imparted knowledge, experience, and valuable lessons that have shaped the author's academic and personal development.

6. The author's parents and family, whose unconditional love, prayers, sacrifices, and unwavering belief have served as a constant source of strength and motivation throughout every stage of the author's education.
7. Carolus Surya Satriyan Putra, for being a dependable companion throughout the thesis process, from the early stages of research to its completion, and for providing encouragement, understanding, and emotional support during both challenging and rewarding moments.
8. Eka Maurita, Hilda Desfianty Arifin, and Oktavia Nur Khasanah, for sharing the experiences, challenges, and memorable moments of university life, as well as for their friendship and support throughout the years of study.
9. Afira Ayu Novanda and Annisah Qotrunnada Purwanto, for their enduring friendship, kindness, and willingness to listen, which provided comfort, motivation, and encouragement during the author's academic journey.
10. All friends from the Informatics Study Program of Universitas Pembangunan Nasional "Veteran" Jawa Timur, along with everyone else who has offered assistance, support, or encouragement, whether directly or indirectly, during the completion of this thesis.

The author acknowledges that this thesis is not without limitations and may still contain shortcomings. Therefore, constructive feedback and suggestions are sincerely welcomed as valuable contributions toward improving the quality of this work.

Surabaya, June 15, 2026

**Author**

## LIST OF CONTENTS

|                                               |             |
|-----------------------------------------------|-------------|
| <b>APPROVAL SHEET .....</b>                   | <b>iii</b>  |
| <b>APPROVAL SHEET .....</b>                   | <b>v</b>    |
| <b>STATEMENT OF ORIGINALITY .....</b>         | <b>vii</b>  |
| <b>ABSTRACT .....</b>                         | <b>ix</b>   |
| <b>ACKNOWLEDGEMENTS.....</b>                  | <b>xi</b>   |
| <b>LIST OF CONTENTS .....</b>                 | <b>xiii</b> |
| <b>LIST OF FIGURES .....</b>                  | <b>xvii</b> |
| <b>LIST OF TABLES .....</b>                   | <b>xix</b>  |
| <b>CHAPTER I INTRODUCTION.....</b>            | <b>1</b>    |
| 1.1. Background .....                         | 1           |
| 1.2. Formulation of the problem.....          | 4           |
| 1.3. Research purposes .....                  | 4           |
| 1.4. Benefits of research.....                | 4           |
| 1.5. Scope of problem .....                   | 5           |
| <b>CHAPTER II LITERATURE REVIEW .....</b>     | <b>7</b>    |
| 2.1. Previous Research .....                  | 7           |
| 2.2. Javanese dialect.....                    | 10          |
| 2.3. Digital Voice Signal Processing.....     | 11          |
| 2.4. Audio Feature Extraction .....           | 12          |
| 2.4.1. <i>Spectrogram</i> .....               | 13          |
| 2.4.2. <i>Mel Spectrogram</i> .....           | 14          |
| 2.4.3. <i>Log Mel-Spectrogram</i> .....       | 15          |
| 2.5. Data Augmentation .....                  | 15          |
| 2.5.1. <i>Noise Injection</i> .....           | 15          |
| 2.5.2. <i>SpecAugment</i> .....               | 16          |
| 2.6. Voice Activity Detection (VAD) .....     | 16          |
| 2.6.1. <i>Silero VAD</i> .....                | 17          |
| 2.7. Deep Learning.....                       | 17          |
| 2.8. Convolutional Neural Network (CNN) ..... | 18          |
| 2.8.1. Feature Learning .....                 | 19          |
| 2.8.2. Classification.....                    | 21          |

|                                     |                                                   |           |
|-------------------------------------|---------------------------------------------------|-----------|
| 2.9.                                | Temporal Convolutional Networks (TCN) .....       | 22        |
| 2.9.1.                              | Causal Convolutions .....                         | 23        |
| 2.9.2.                              | Dilated Convolutions.....                         | 23        |
| 2.9.3.                              | Residual Blocks.....                              | 25        |
| 2.10.                               | Ensemble Learning.....                            | 26        |
| 2.10.1.                             | Soft Voting .....                                 | 26        |
| 2.11.                               | Interface Development Framework (Streamlit) ..... | 27        |
| 2.12.                               | Performance Evaluation Metrics .....              | 27        |
| 2.12.1.                             | Confusion Matrix .....                            | 28        |
| 2.12.2.                             | Accuracy.....                                     | 28        |
| 2.12.3.                             | Precision .....                                   | 29        |
| 2.12.4.                             | Recall.....                                       | 29        |
| 2.12.5.                             | F1-Score .....                                    | 30        |
| <b>CHAPTER III METHODOLOGY.....</b> |                                                   | <b>31</b> |
| 3.1.                                | Literature Study.....                             | 31        |
| 3.2.                                | Research Stages.....                              | 31        |
| 3.3.                                | Data collection.....                              | 32        |
| 3.3.1.                              | Data Selection Criteria .....                     | 33        |
| 3.3.2.                              | Additional Data Source Validation Procedure ..... | 34        |
| 3.3.3.                              | Dialect Grouping Strategy.....                    | 35        |
| 3.3.4.                              | Data Distribution .....                           | 36        |
| 3.4.                                | Data Labeling .....                               | 36        |
| 3.5.                                | Data Preprocessing.....                           | 37        |
| 3.5.1.                              | Frequency Standardization(Resampling) .....       | 37        |
| 3.5.2.                              | Voice Activity Detection.....                     | 39        |
| 3.5.3.                              | Audio Segmentation.....                           | 40        |
| 3.6.                                | Data Augmentation (Noise Injection) .....         | 41        |
| 3.7.                                | Data Splitting.....                               | 42        |
| 3.8.                                | Feature Extraction .....                          | 43        |
| 3.9.                                | Feature Normalization.....                        | 46        |
| 3.10.                               | Data Augmentation (SpecAugment) .....             | 48        |
| 3.11.                               | Designing a Hybrid CNN-TCN Model .....            | 49        |

|                     |                                                     |            |
|---------------------|-----------------------------------------------------|------------|
| 3.11.1.             | Front-End: Convolutional Neural Network (CNN) ..... | 50         |
| 3.11.2.             | Back-End: Temporal Convolutional Network (TCN)..... | 53         |
| 3.11.3.             | Classification Layer.....                           | 55         |
| 3.12.               | Testing Scenario.....                               | 56         |
| 3.13.               | Prediction Aggregation and Model Evaluation .....   | 58         |
| 3.14.               | System Implementation.....                          | 59         |
| <b>CHAPTER IV</b>   | <b>RESULTS AND DISCUSSION.....</b>                  | <b>63</b>  |
| 4.1.                | Data Collection and Labeling.....                   | 63         |
| 4.2.                | System Implementation.....                          | 64         |
| 4.2.1.              | Data Pre-processing.....                            | 65         |
| 4.2.2.              | Data Augmentation (Noise Injection) .....           | 70         |
| 4.2.3.              | Data Splitting.....                                 | 73         |
| 4.2.4.              | Feature Extraction .....                            | 74         |
| 4.2.5.              | Feature Normalization.....                          | 76         |
| 4.2.6.              | Data Augmentation(SpecAugment) .....                | 79         |
| 4.2.7.              | Architectural Design Model .....                    | 81         |
| 4.3.                | Model Training.....                                 | 89         |
| 4.4.                | Model Testing .....                                 | 91         |
| 4.4.1.              | Ablation Study Testing.....                         | 91         |
| 4.4.2.              | Hyperparameter Optimization Testing.....            | 100        |
| 4.4.3.              | Robustness Testing on Mixed Data.....               | 108        |
| 4.5.                | Results Analysis .....                              | 110        |
| 4.6.                | Web Implementation and Interface .....              | 112        |
| <b>CHAPTER V</b>    | <b>CONCLUSION AND SUGGESTIONS.....</b>              | <b>115</b> |
| 5.1.                | Conclusion.....                                     | 115        |
| 5.2.                | Suggestions.....                                    | 115        |
| <b>BIBLIOGRAPHY</b> | <b>.....</b>                                        | <b>117</b> |
| <b>APPENDICES</b>   | <b>.....</b>                                        | <b>127</b> |

## LIST OF FIGURES

|                                                                                                     |    |
|-----------------------------------------------------------------------------------------------------|----|
| Figure 2.1 Block Diagram of Analog-to-Digital Conversion (ADC) Process .....                        | 11 |
| Figure 2.2 Deep Neural Network Architecture .....                                                   | 18 |
| Figure 2.3 CNN Architecture .....                                                                   | 18 |
| Figure 2.4 Pooling Layer.....                                                                       | 21 |
| Figure 2.5 Visualization of Causal Convolution and Non-causal Convolution....                       | 23 |
| Figure 2.6 Visualization of Dilated Causal Convolution and Causal Convolution<br>.....              | 24 |
| Figure 2.7 Residual Blocks Visualization .....                                                      | 25 |
| Figure 2.8 Streamlit Logo .....                                                                     | 27 |
| Figure 3.1 Research Stages Flow .....                                                               | 32 |
| Figure 3.2 TLA Digital Repository Interface.....                                                    | 33 |
| Figure 3.3 Data Pre-processing Flow .....                                                           | 37 |
| Figure 3.4 Sampling Rate Distribution .....                                                         | 38 |
| Figure 3.5 Comparison of Raw and Resampled Waveforms .....                                          | 38 |
| Figure 3.6 Audio Segmentation Illustration (Duration 10s, Overlap 30%) .....                        | 40 |
| Figure 3.7 Transforming an Audio Signal into a Log-Mel Spectrogram.....                             | 44 |
| Figure 3.8 Feature Data Distribution Comparison .....                                               | 48 |
| Figure 3.9 Proposed Architecture Diagram of Hybrid CNN-TCN .....                                    | 50 |
| Figure 3.10 CNN Convolutional Block Architecture Details .....                                      | 51 |
| Figure 3.11 TCN Residual Block Architecture Details.....                                            | 54 |
| Figure 3.12 Website Appearance Wireframe .....                                                      | 60 |
| Figure 4.1 Visualization Comparison of Sample Rate Distribution Before and After<br>Resampling..... | 66 |
| Figure 4.2 Recapitulation of Duration per Dialect After the VAD Process.....                        | 68 |
| Figure 4.3 Recapitulation of the Number of Segments in Each Dialect .....                           | 69 |
| Figure 4.4 Data Splitting Validation Process Output on Fold .....                                   | 74 |
| Figure 4.5 Loss and Accuracy Graph Scenario 1 (Fold 3).....                                         | 92 |
| Figure 4.6 Classification Report Segment Scenario 1 .....                                           | 92 |
| Figure 4.7 Classification Report Soft Voting Scenario 1 .....                                       | 93 |
| Figure 4.8 Audio Level Confusion Matrix in Scenario 1.....                                          | 94 |

|                                                                         |     |
|-------------------------------------------------------------------------|-----|
| Figure 4.9 Loss and Accuracy Graph Scenario 2 (Fold 3).....             | 95  |
| Figure 4.10 Classification Report Segment Scenario 2 .....              | 96  |
| Figure 4.11 Classification Report Soft Voting Scenario 2.....           | 96  |
| Figure 4.12 Confusion Matrix Scenario 2.....                            | 97  |
| Figure 4.13 Loss and Accuracy Graph Scenario 3 (Fold 2).....            | 98  |
| Figure 4.14 Classification Report Soft Voting Scenario 3.....           | 98  |
| Figure 4.15 Classification Report Segment Scenario 3 .....              | 99  |
| Figure 4.16 Confusion Matrix Scenario 3.....                            | 99  |
| Figure 4.17 Loss and Accuracy Graph Scenario 4 (Fold 4).....            | 100 |
| Figure 4.18 Classification Report Segment Scenario 4 .....              | 101 |
| Figure 4.19 Classification Soft Voting Report Scenario 4.....           | 101 |
| Figure 4.20 Confusion Matrix Scenario 4.....                            | 102 |
| Figure 4.21 Loss and Accuracy Graph for Scenario 5 (Fold 5) .....       | 103 |
| Figure 4.22 Classification Report Scenario Segment 5 .....              | 104 |
| Figure 4.23 Classification Soft Voting Report Scenario 5.....           | 104 |
| Figure 4.24 Confusion Matrix Scenario 5.....                            | 105 |
| Figure 4.25 Loss and Accuracy Graph Scenario 6 (Fold 5).....            | 106 |
| Figure 4.26 Classification Report Scenario Segment 6 .....              | 106 |
| Figure 4.27 Classification Report Soft Voting Scenario 6.....           | 107 |
| Figure 4.28 Confusion Matrix Scenario 6.....                            | 107 |
| Figure 4.29 Javanese Dialect Identification Web Interface Display ..... | 113 |

## LIST OF TABLES

|                                                                                       |     |
|---------------------------------------------------------------------------------------|-----|
| Table 2.1 Previous Research .....                                                     | 10  |
| Table 3.1 YouTube Channel Source Details .....                                        | 34  |
| Table 3.2 Final Dialect Class and Basis for Territorial Determination.....            | 35  |
| Table 3.3 Raw Data Distribution per Dialect Class .....                               | 36  |
| Table 3.4 Raw Log-Mel Spectrogram Data Sample (Bin 10, Frame 156) .....               | 47  |
| Table 3.5 Testing Scenario.....                                                       | 57  |
| Table 4.1 Audio File Naming Standardization Results.....                              | 64  |
| Table 4.2 Audio Resampling Results.....                                               | 66  |
| Table 4.3 Final Distribution of Audio Files and Segments per Dialect Class .....      | 70  |
| Table 4.4 Waveform Visualization Before and After Augmentation .....                  | 72  |
| Table 4.5 Waveform Visualization and Log Mel Spectrogram .....                        | 75  |
| Table 4.6 Visualization of SpecAugment Transformation on Log-Mel Spectrogram<br>..... | 80  |
| Table 4.7 Comparison of the Results of the Robustness Test of All Scenarios...        | 108 |
| Table 4.8 Summary of Test Scenario Evaluation Results (5-Fold Average).....           | 110 |