UNDERGRADUATE THESIS

HANDWRITTEN JAVANESE SCRIPT TEXT RECOGNITION USING CNN BILSTM AND CONNECTIONIST TEMPORAL CLASSIFICATION

ADE FATHONI PRASTYA
NPM 22081010204

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

HANDWRITTEN JAVANESE SCRIPT TEXT RECOGNITION USING CNN BILSTM AND CONNECTIONIST TEMPORAL CLASSIFICATION

By:
ADE FATHONI PRASTYA
NPM. 22081010204

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 May 25, 2026:

Approved,

Fetty Tri Anggraeny, S.Kom., M.Kom. NIP. 19820211 2021212 005	.....	**(Advisor I)**
Eva Yulia Puspaningrum, S.Kom., M.Kom. NIP. 19890705 2021212 002	.....	**(Advisor II)**
Dr. Faisal Muttaqin, S.Kom., MT. NIP. 19851231 202121 1 009	.....	**(Head Assessor)**
Muhammad Muharrom Al Haromainy, S.Kom., M.Kom. NIP. 19950601 202203 1 006	.....	**(Assessor I)**

Acknowledge by,

Dean of the Faculty of Computer Science

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

APPROVAL SHEET

HANDWRITTEN JAVANESE SCRIPT TEXT RECOGNITION USING CNN BILSTM AND CONNECTIONIST TEMPORAL CLASSIFICATION

By:
ADE FATHONI PRASTYA
NPM. 22081010204

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 202521 2 029

STATEMENT OF ORIGINALITY

I am the undersigned:

Student Name	:	ADE FATHONI PRASTYA
NPM	:	22081010204
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, May 25, 2026
Author,

ADE FATHONI PRASTYA
NPM. 22081010204

ABSTRACT

Student Name / NPM	:	Ade Fathoni Prastya / 22081010204
Thesis Title	:	Handwritten Javanese Script Text Recognition Using CNN BiLSTM and Connectionist Temporal Classification
Supervisor	:	1. Fetty Tri Anggraeny, S.Kom., M.Kom.
	:	2. Eva Yulia Puspaningrum, S.Kom., M.Kom.

The Nglegena Javanese script is a traditional writing system with high historical and cultural value, but faces the threat of extinction due to the low literacy skills of the community. Based on previous research, 81.4% of students are unable to read and 79% are unable to write Javanese script. On the other hand, more than 19,000 Javanese script manuscripts stored in various collections have not been adequately inventoried, so an automatic text recognition system is needed that can support digitization and preservation efforts. This study aims to develop an end-to-end handwritten text recognition model for the Nglegena Javanese script based on Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and Connectionist Temporal Classification (CTC) architecture without requiring explicit character segmentation. The dataset used consists of 1,200 synthetic images resulting from rendering Javanese script fonts, with 1,000 images as training data and 200 images as validation data. In addition, 300 real handwritten images from six participants were used, divided into 100 images as validation data and 200 images as test data. This study systematically explored architectural variations by varying the number of CNN layers from 3 to 7 layers and BiLSTM layers from 1 to 3 layers, using Character Error Rate (CER) and Exact Match (EM) as evaluation metrics. The experimental results showed that the optimal configuration was achieved by the 5-CNN and 2-BiLSTM architectures with a CER value of 0.068 and an EM accuracy of 0.795. Fine architectures (3-4 CNN layers) indicated underfitting due to limited feature capacity, while deeper architectures (6-7 CNN layers and 3 BiLSTM layers) showed performance degradation and instability due to overfitting. The effectiveness of the CNN-BiLSTM-CTC architecture synergy was proven through the model's high generalization ability on varied real handwritten test data. The entire system is then implemented in the form of an API and an interactive website interface to support accessibility of Nglegena Javanese script recognition by other systems and the wider community.

Keywords: Nglegena Javanese Script, Handwritten Text Recognition, CNN, BiLSTM, CTC, Character Error Rate, Exact Match.

ACKNOWLEDGEMENTS

Praise and thanks be to Allah SWT for all His mercy, guidance, and blessings, which have enabled the author to successfully complete this thesis titled **"Handwritten Javanese Script Text Recognition Using CNN BiLSTM and Connectionist Temporal Classification."**

The author would like to express gratitude to Ms. Fetty Tri Anggraeny, S.Kom., M.Kom., as the primary advisor, and Ms. Eva Yulia Puspaningrum, S.Kom., M.Kom., as the co-advisor, for taking the time to provide invaluable guidance, advice, and motivation to the author. The author has also received significant support from various parties, whether in the form of moral, spiritual, or material assistance. For this, the author would like to thank:

1. Prof. Dr. Ir. Novirina Hendrasarie, M.T., Dean of the Faculty of Computer Science at the Veteran National Development University of East Java.
2. Dr. Intan Yuniar Purbasari, S.Kom., M.Sc., Coordinator of the Informatics Study Program at the Faculty of Computer Science, Veteran National Development University of East Java.
3. Dr. Faisal Muttaqin, S.Kom., M.T., and Muhammad Muharrom Al Haromainy, S.Kom., M.Kom., as Seminar Proposal Examiners who provided constructive suggestions, feedback, and criticism.
4. All faculty members of the Computer Science Program, Faculty of Computer Science, Veteran National Development University of East Java, who equipped the author with knowledge during the course of study.
5. All classmates from the 2022 cohort of the Informatics Program for their camaraderie and support.
6. Edmin Adi Prasetyo, Pangelora Ridho August Susilo, Rizky Ramadhani, Fernando Putra Pratama, and Achmad Ardian, as participants in the compilation of the Nglegena Javanese script dataset.
7. My parents and beloved family, for their unceasing prayers, love, and both moral and material support.

The author acknowledges that this thesis is far from perfect. Therefore, the author welcomes constructive criticism and suggestions. Finally, I hope this thesis will

contribute to the advancement of knowledge, particularly in the field of computer science.

Surabaya, May 25th 2026

Author

ADE FATHONI PRASTYA

NPM. 22081010204

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SOURCE CODES