# MAHATMA GANDHI INSTITUTE OF TECHNOLOGY

(Estd in 1997 by Chaitanya Bharathi Educational Society)

(Affiliated to JNTU, Hyderabad; Accredited by NBA, AICTE-New Delhi)

Kokapet(village), Rajendra Nagar(Mandal), Ranga Reddy(Dist.),

Chaitanya Bharathi P.O., Hyderabad-500075.

## CERTIFICATE

This is to certify that the project work entitled, ***TELUGU WORD RECOGNITION USING CNN*** being submitted by **P.SAI KRISHNA (Roll No.15261A0539) & N.CHANDRA SHEKAR (Roll No. 16265A0510)** in partial fulfillment of the requirements for the Subject Major Project (A80088) in **COMPUTER SCIENCE AND ENGINEERING** is a record of bonafide work carried out by them. The results of investigations enclosed in this report have been verified and found satisfactory.

**PROJECT GUIDES** **HEAD OF THE DEPARTMENT**

**Dr. V.SUBBA RAMAIAH**
**Assistant Professor**

**Dr. C.R.K. REDDY**
**Professor**
**Department of CSE**

**Mr. D.KOTESWARA RAO**
**Assistant Professor**

**EXTERNAL EXAMINER**

# DECLERATION

This is to certify that the work reported in this project titled "**TELUGU WORD RECOGNITION USING CNN**" is a record of work done by us in the department of Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Hyderabad. No part of the work is copied from books/journals/internet and wherever the portion is taken, the same has been duly referred in the text. The report is based on the project work done entirely by us and not copied from any other source.

**Place**: Gandipet
**Date**: 20/04/2019

**P.SAI KRISHNA**
**15261A0539**
**N.CHANDRA SHEKAR**
**16265A0510**

# ACKNOWLEDGEMENT

We are also thankful to our Principal **Prof. K. JAYA SANKAR** for providing excellent infrastructure and a nice atmosphere for completing this project successfully.

Our sincere thanks to **Dr. C. R. K. REDDY** Head of the Department, CSE, who inspired us with his valuable suggestions and advice throughout graduation studies at the college and also during our project work.

We would like to express our sincere gratitude and indebtedness to our project course guides **Dr. V.SUBBA RAMAIAH** and **Mr. D.KOTESWARA RAO** for their valuable suggestions and interest throughout the course of this project.

We convey our heartfelt thanks to the lab staff for allowing us to use the required equipment whenever needed.

Finally, we would like to take this opportunity to thank our families for their support through the work. We sincerely acknowledge and thank all those who gave directly or indirectly their support in completion of this work.

**P.SAI KRISHNA (Roll No.15261A0539)**

**N.CHANDRA SHEKAR (Roll No.16265A0510)**

# TABLE OF CONTENTS

# LIST OF FIGURES

# ABSTRACT

Page segmentation is the process by which a scanned page is divided into columns and blocks which are then classified as graphics, or text based on some properties like width, height and area of the connected components. In this project Run Length Smearing algorithm (RLSA) is implemented for page segmentation and Convolutional neural network is used words recognition.

The main aim of this project is to apply the segmentation algorithm for Telugu Document images to obtain regions of interest like graphics and text. On performing segmentation on Telugu document images containing both text and graphics, the segmentation algorithm separates the text and graphics regions. The next step is to use the words images along with their labels for training and text recognition by using a convolutional neural network.