**DESIGN AND DEVELOPMENT OF HANDWRITTEN TEXT RECOGNITION THROUGH NEURAL NETWORK AND MACHINE LEARNING TECHNIQUES**

##### A PROJECT REPORT

###### *Submitted by*

##### JISHNU CK (721415104015)

*in partial fulfillment for the award of the degree*

*of*

**BACHELOR OF ENGINEERING**

*in*

# COMPUTER SCIENCE AND ENGINEERING



NEHRU INSTITUE OF ENGINEERING AND TECHNOLOGY COIMBATORE

ANNA UNIVERSITY :: CHENNAI 600 025

##### MARCH 2019

ANNA UNIVERSITY :: CHENNAI 600 025

# ****BONAFIDE CERTIFICATE****

Certified that this project report **“DESIGN AND DEVELOPMENT OF HANDWRITTEN TEXT RECOGNITION THROUGH NEURAL NETWORK AND MACHINE LEARNING TECHNIQUES”** is the bonafide work of **JISHNU CK (721415104015)**  who carried out the project work under my supervision.

|  |  |
| --- | --- |
| **SIGNATURE**  **DR. A. SURESH**  **HEAD OF THE DEPARTMENT**  Professor  Department of Computer Science & Engineering  Nehru Institute of Engineering & Technology  Coimbatore - 641105 | **SIGNATURE**  **DR.S.SUBASREE**  **SUPERVISOR**  Professor  Department of Computer Science & Engineering  Nehru Institute of Engineering & Technology  Coimbatore - 641105 |

Submitted for the Anna University project viva – voice examination held on ……………… at Nehru Institute of Engineering and Technology.

|  |  |
| --- | --- |
| **INTERNAL EXAMINER** | **EXTERNAL EXAMINER** |

# ****ACKNOWLEDGEMENT****

At this pleasing moment of successful completion of my project, I thank God Almighty for his blessings.

I convey our sincere thanks to our respected Founder Chairman Late **Shri. P. K.** **DAS**, Chairman and Managing Trustee **Adv. Dr. P KRISHNADAS** and CEO & Secretary **Dr. P. KRISHNAKUMAR** who provided the facilities for us.

I express our deep gratitude to our beloved and honorable Principal, **Dr. P. MANIIARASAN** and Dean, Academic Affairs **Dr.N.K.Sakthivel** for their valuable support.

I owe grateful and significant thanks to our Head of Department, Computer Science and Engineering, **Dr. A. SURESH** who has given constant support and encouragement throughout the completion of our course and our project work.

I thank my Internal Guide **DR.S.SUBASREE,** Professor-CSE, who rendered us all possible help along with her valuable suggestions and constructive guidance for successful completion of our project.

I would like to express our special gratitude to our Project Coordinator **Ms. S. JOTHIMANI**, Assistant Professor-CSE, for providing necessary information regarding the project and also for her support in completing the project.

I am thankful for the support and guidance from all the **faculty members of Computer Science and Engineering Department** and sincere thanks to our supporting staff for their timely assistance.

Finally I would like to express our gratitude towards parents for their motivation and support which helped us in the completion of this project.

# ABSTRACT

Humans have the ability to recognize characters. For example, human can distinguish between different characters and recognize them easily as an A or a B and so on. Therefore, project is intended to develop a neural network system that is able to perform character recognition, particularly English alphabets. Neural network is a system inspired by human brain function; consists of neurons connected in parallel with the ability to learn. A basic design of neural network has input layer, hidden layer, and output layer. The use of neural network can improve the quality of recognition while achieving good performance. A total of 650 samples characters are used with 25 samples of each character. The performance of evaluation is divided to 80% of training and 20% for testing. Scaled conjugate gradient training function is used as this function can perform faster in pattern recognition as well as its small memory requirement. Two training methods are used. The first one is the Gradient Technique with 39 neurons in hidden layer. The second training method is Geometric Feature Extraction with 35 neurons in hidden layer. Gradient Technique and Geometric Feature Extraction; both show an excellent recognition rate of 94.6% and 94.3% respectively. The output of recognized characters is shown in a .txt file.

Chapter 1

INTRODUCTION TO HANDWRITTEN CHARACTER RECOGNITION

* 1. Introduction

The development of handwriting recognition systems began in the 1950s when there were human operators whose job was to convert data from various documents into electronic format, making the process quite long and often affected by errors. Automatic text recognition aims at limiting these errors by using image preprocessing techniques that bring increased speed and precision to the entire recognition process. Handwriting recognition has been one of the most fascinating and challenging research areas in field of image processing and pattern recognition in the recent years. It contributes immensely to the advancement of automation process and improves the interface between man and machine in numerous applications. Optical character recognition is a field of study than can encompass many different solving techniques. Neural networks (Sandhu & Leon, 2009), support vector machines and statistical classifiers seem to be the preferred solutions to the problem due to their proven accuracy in classifying new data.

he Optical Character Recognizer actually is a convertor which translates handwritten text images to a machine based text. In general, handwriting recognition is classified into two types as off-line and on-line. In the off-line recognition, the writing is usually capture optically by a scanner and the completed writing is available as an image. In other words, Offline Handwritten Text is when hand written text is scanned by a scanner into a digital format. But, in the on-line system the two dimensional coordinates of successive points are represented as a function of time and the order of strokes made by the writer. In other words, X-Y coordinates are given as a result that tells the location of the pen and the force applied by the user during writing and speed too. Online Handwritten Text is written by a stylus on a tablet. There is also a third method which is not as famous as the first two methods mentioned above in which laser, inkjet devices, can be used for obtaining machine printed text.

There is extensive work in the field of handwriting recognition, and a number of reviews exist. The on-line methods have been shown to be superior to their off-line counterparts in recognizing handwritten characters due to the temporal information available with the former [3] [4]. However, several applications including mail sorting, bank processing, document reading and postal address recognition require off-line handwriting recognition systems. Moreover, in the off-line systems, the neural networks and support vector machines have been successfully used to yield comparably high recognition accuracy levels. As a result, the off-line handwriting recognition continues to be an active area for research towards exploring the newer techniques that would improve recognition accuracy. Therefore, for this report, I have decided to work on an off-line handwritten alphabetical character recognition system using CNN,RNN and CTC.

Artificial Neural Network (ANN) is a computing model of brain, having paralleled distributed processing elements that are learned by adjusting the connected weights between the neurons. Due to its flexibility and strength, it has been now broadly used in different fields such as pattern recognition, decision-making optimization, market analysis, robot intelligence [7]. ANN can be more remarkable as computational processors for different tasks like data compression, classification, combinatorial optimization problem solving, pattern recognition etc. [8]. ANN has many advantages over the other classical methods. While having the computational complexity, ANN offered many advantages in pattern recognition adapting a very little context of human intelligence [9]. In the off-line recognition system, the neural networks have emerged as the fast and reliable tools for classification towards achieving high recognition accuracy [10]. Classification techniques have been applied to handwritten character recognition since the 1990s. These methods include statistical methods based on Bayes decision rule, Artificial Neural Networks (ANNs), Kernel Methods including Support Vector Machines (SVM) and multiple classifier combination

* 1. Objective