**Design Report for**

# Continues Integration Pipeline Implementation for Tech11Software

*Submitted By*


***Aswin G Sugunan*** *(13)*

***Jefin Jacob*** *(4)*

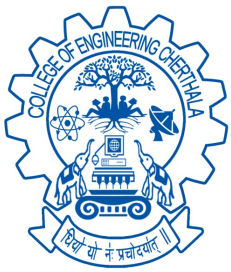
***Nitin Suresh*** *(23)*

***Vishnu Bose*** *(39)*


***7th Semester***

*Under the guidance of*

***Mrs. Greeshma N Gopal***




**JULY 2016**


**Department of Computer Science and Engineering**
**College of Engineering,Cherthala**
Pallippuram P O,Alappuzha-688541
Phone: 0478 2553416, Fax: 0478 2552714
http://www.cectl.ac.in

# Contents

# List of Figures

# Chapter 1

# INTRODUCTION

Software development, as we know it today, is a demanding area of business with its fast-changing customer requirements, pressures of an ever shorter timeto- market, and unpredictability of market. With the shift towards modern continuous deployment pipelines, releasing new software versions early and often has become a concrete option also for an ever growing number of practitioners.

Continuous delivery is a software development practice where new features are made available to end users as soon as they have been implemented and tested. In such a setting, a key technical piece of infrastructure is the development pipeline that consists of various tools and databases, where features ow from development to deployment and then further to use.

## 1.1 Purpose

The purpose of the design document is to describe the behavior of the proposed CI framework system. Requirements Specification defines and describes the operations, interfaces, performance, and quality assurance requirements of the proposed system. The document describes the design constraints that are to be considered when the system is to be designed, and other factors necessary to provide a complete. The Design report analyse the SRS report and converted to implementation form by means od necessary diagrams (DFD, sequence diagram,structure Diagram, ER diagrams) and also mention the system modules and briefly their operations

## 1.2 Product Scope

The objective of the project is to put in place a Continues Integration framework for product development activities of Tech11 Software.This would enable the Tech11 team to rapidly bring a product change or feature to production gaining market advantage. This activities of this project will involve accessing different CI integration approaches and solutions available,identify the feasibility of those solution by doing POCs and demos,fine tune the final solution and set up the CI infrastructures,educate the developers on CI culture.

# Chapter 2

# OVERALL DESCRIPTION

## 2.1 Product Perspective

Continuous Integration is based on continuous performance of acts of integration of source code, testing, building and deployment in response to each change to the source code of the project submitted by the developer and for the use of tools for support of the development and testing by compliance with the established procedure automatically. The proposed system directs the user (developers) for time and cost effective production and solves majority of the Integration hell problem.

Integration Hell refers to the point in production when members on a delivery team integrate their individual code. In traditional software development environments, this integration process is rarely smooth and seamless, instead resulting in hours or perhaps days of fixing the code so that it can finally integrate. Continuous Integration (CI) aims to avoid this completely by enabling and encouraging team members to integrate frequently (e.g., hourly, or at least daily).

### 2.1.1 Proposed System

The Proposed system is to help software developers to ensure new features are made available as soon as the program has been implemented and tested. This product also helps in reducing the time needed to develop a software and also acts a guideline for future software developments.

## 2.2 Product Function

- Data Acquisition

  This is the stage in which data are collected as part of the recognition process. The data may be captured scanning the image after the writing process is over.It is user's responsibility to take better scanned image during data acquisition process

- Pre Processing

  Pre Processing is an important step in character recognition process because of the variations in the writing style among different users and the existence of huge amount of noise in the images after scanning.

- Segmentation

  Segmentation is an operation that isolates individual characters from the handwritten text. It is done using projection profile analysis and connected component labeling.A popular technique used for line segmentation is horizontal projection profile method in which peak-valley points are identified and is used for line separation. The horizontal projection will have separated peaks and valleys if the lines are well separated. These peaks can be used to find out the boundaries between lines and so each line is extracted. Word segmentation is done by applying vertical projection profile method on the separated lines. Here the peaks and valleys are identified and words are separated by looking at the minimal in the vertical profile. Finally, the characters are isolated from these words using connected component labeling. Connected components labeling groups pixels of an image into components based on pixel connectivity. This technique assigns to each connected component of a binary image a distinct label.

- Feature Extraction

  Feature extraction is the process of extracting relevant features of the characters to form feature vectors which are used by classifiers for the recognition process. The feature extraction methods for handwritten character recognition can be classified into three: Sta-

tistical, Structural and Hybrid techniques. Statistical approaches use quantitative methods for extracting the features. Geometrical moments, projection histograms, direction histograms, crossing points etc. are used as features here. Structural approaches use qualitative measurements for feature extraction. These features are based on topological and geometrical properties of the character, like strokes, loops, end points, intersection points, etc. Hybrid approaches combines the features of these two techniques.

- Classification

  Classification is the final phase of character recognition, which is done by assigning labels to character images based on the features extracted. Bayesian classifier, Binary tree classifier, Nearest Neighbor classifier, Neural networks, MQDF and Support Vector Machines are some of the classifiers that are used for this purpose.In this project we are using Neural networks for doing the classification process.

- Post Processing

  Post-processing involves steps to be taken after classification process is completed. It may include steps like representing the output in Unicode format, error correction and the disambiguation of confusing character pairs. Linguistic rules can also be applied to further improve recognition rates. These linguistic rules are specific to a language.

## 2.3 Operating Environment

The system is expected to be operated in Linux as well as in windows with the support of respective JRE (Java Runtime Environment). This system based project is completely platform independent. The most important requirement is the internet connection. This tool is coded using JDK 1.6.

# Chapter 3

# PROJECT REQUIREMENTS

This system based software can be used in any operating system such as Microsoft Windows, Linux or any kind of user application interface, since it is platform independent.

## 3.1 User Interface

Interface hardware shall be encapsulated in a set of classes that isolate hardware specifications from the rest of the software. In particular, interfaces for specific hardware boards shall be implemented as derived classes of an abstract class.

## 3.2 Hardware Requirements

| | | |
|---|---|---|
| Processor | : | Min 2GHZ |
| RAM | : | Min 2GB |
| Hard disk | : | Min 200GB |

## 3.3 Network Requirements

- Systems need high speed internet connection.

## 3.4 Software Requirements

- Java, Net Beans IDE
- Internet explorer or any other supported browsers

## 3.5 Gantt Chart

Gantt Chart provides an approximation about the completion of project.It gives a rough description of the completion time,testing time etc.

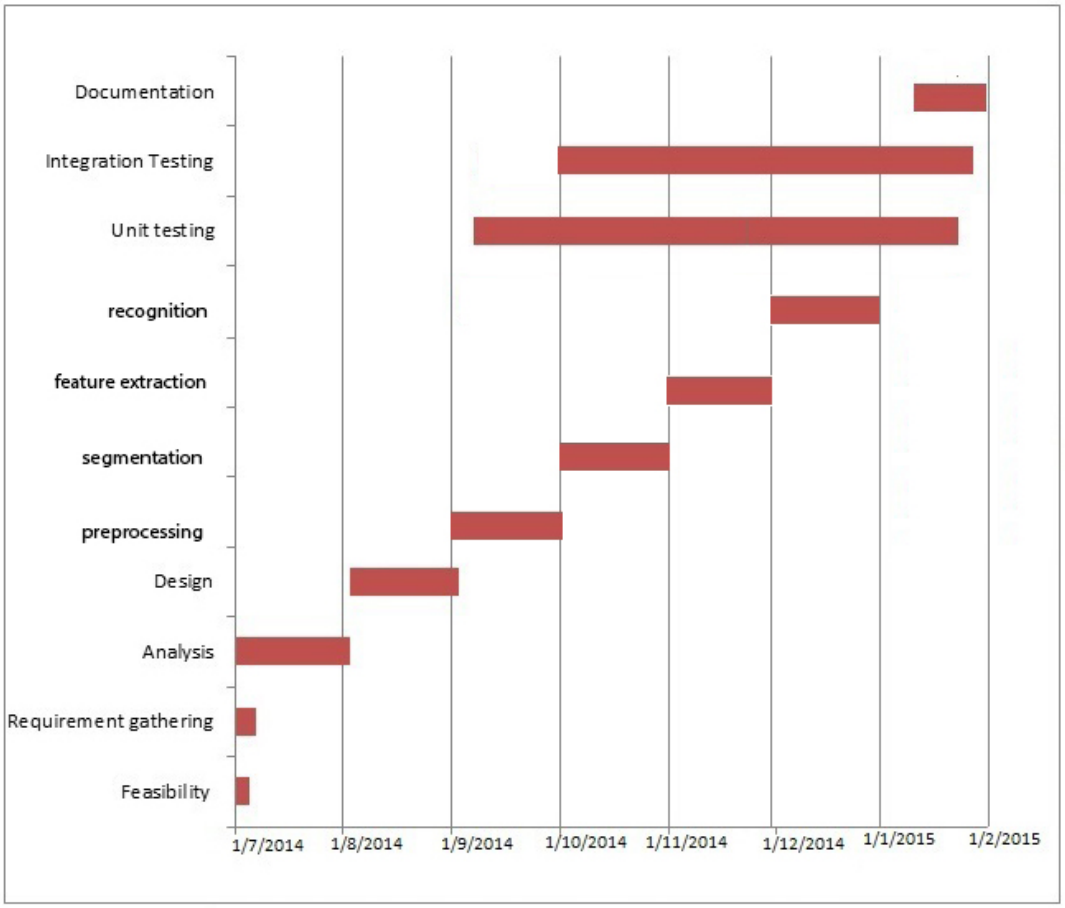



Fig. 3.1: Gantt Chart

# Chapter 4

# SYSTEM DESIGN

## 4.1 Modules

Based on the project aspect the project is divided into 4 modules:

- Input
- Feature Extraction
- Recognition
- Output

### 4.1.1 Input

Input to the software can be of any form. It can either be a scanned image of a handwritten document or a printed text. We can also download the image using the Download option on the GUI. The input has a lower cut off for its quality.

### 4.1.2 Feature Extraction

Feature Extraction of input is the first phase of processing. Each and every character should be separated for its easy recognition. And the separated characters will be fit into a matrix of standard size.

### 4.1.3 Recognition

As soon as the characters are fit into the matrix, the occupied cells should be given value 1 and 0 otherwise. Each and every character is given a certain value range for matrix with respect to the occupied cells.

### 4.1.4 Output

Output can be obtained either as a text file or an audio file as we wish.

## 4.2 Data Flow Diagram

The data flow diagram (DFD) is used for classifying system requirements to major transformation that will become programs in system design. This is starting point of the design phase that functionally decomposes the required specifications down to the lower level of details. It consists of a series of bubbles joint together by lines. Bubbles: Represent the data transformations. Lines: Represent the logic flow of data. Data can trigger events and can be processed to useful information. Systems analysis recognizes the central goal of data in organizations. This dataflow analysis tells a great deal about how organization objectives are accomplished. Dataflow analysis studies the use of data in each activity. It documents these finding in the DFDs. Dataflow analysis give the activities of a system from the view point of data where it originates , how they are used or hanged or where they go, including the stops along the way from their destination. The components of dataflow strategy span both the requirements and systems design. The first part is called dataflow analysis.

### 4.2.1 Level 0 DFD

Level 0 DFD gives a simple information about the overall structure.

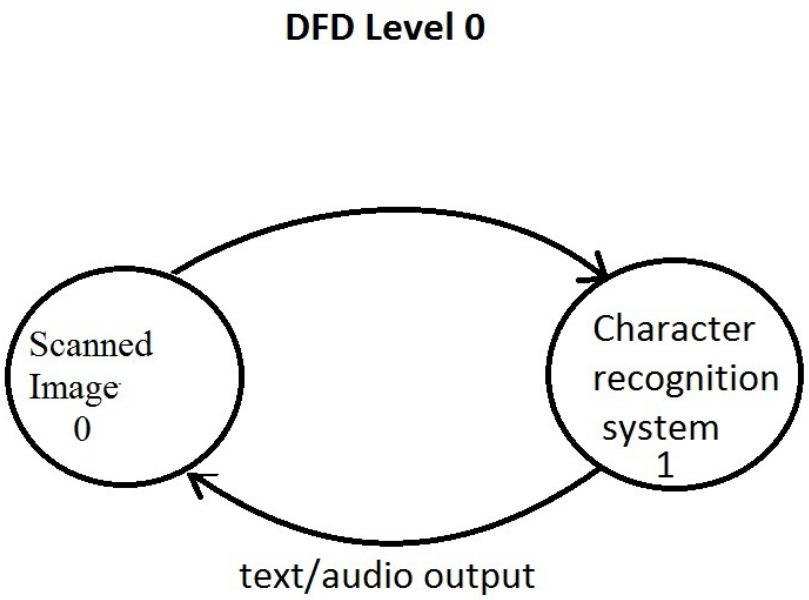



Fig. 4.1: Level 0 DFD

### 4.2.2 Level 1 DFD

The level 1 DFD gives a basic structure of the project indicating the five modules needed to execute the project.

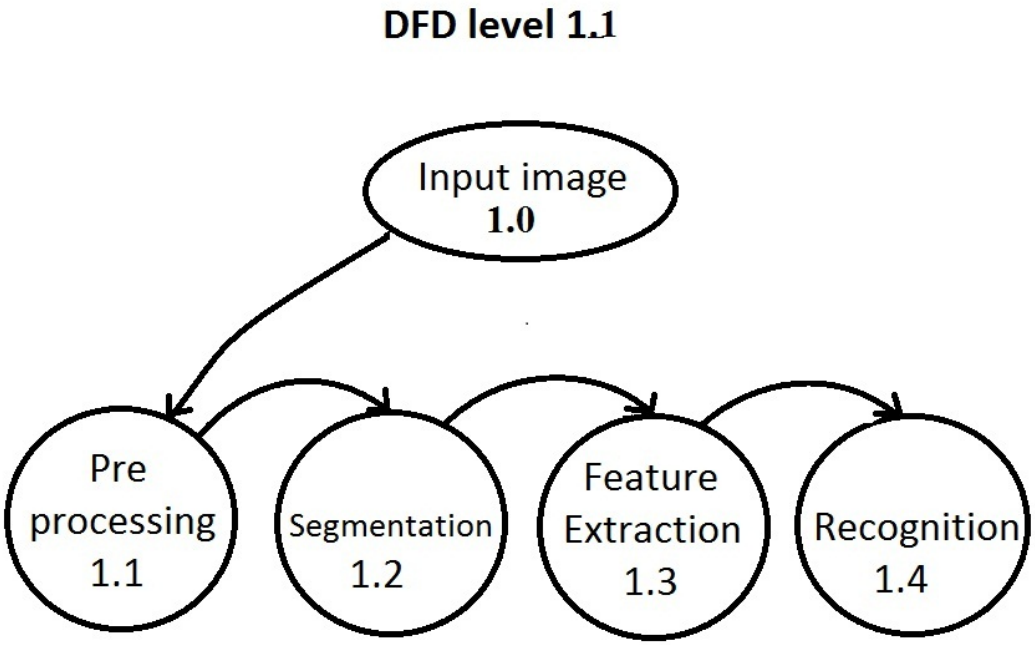



Fig. 4.2: Level 1.1 DFD

### 4.2.3 Level 2 DFD

The level 2 DFD gives a much more advanced idea about the execution. Each of these sections perform unique functions and these are combined together to yield the final product.

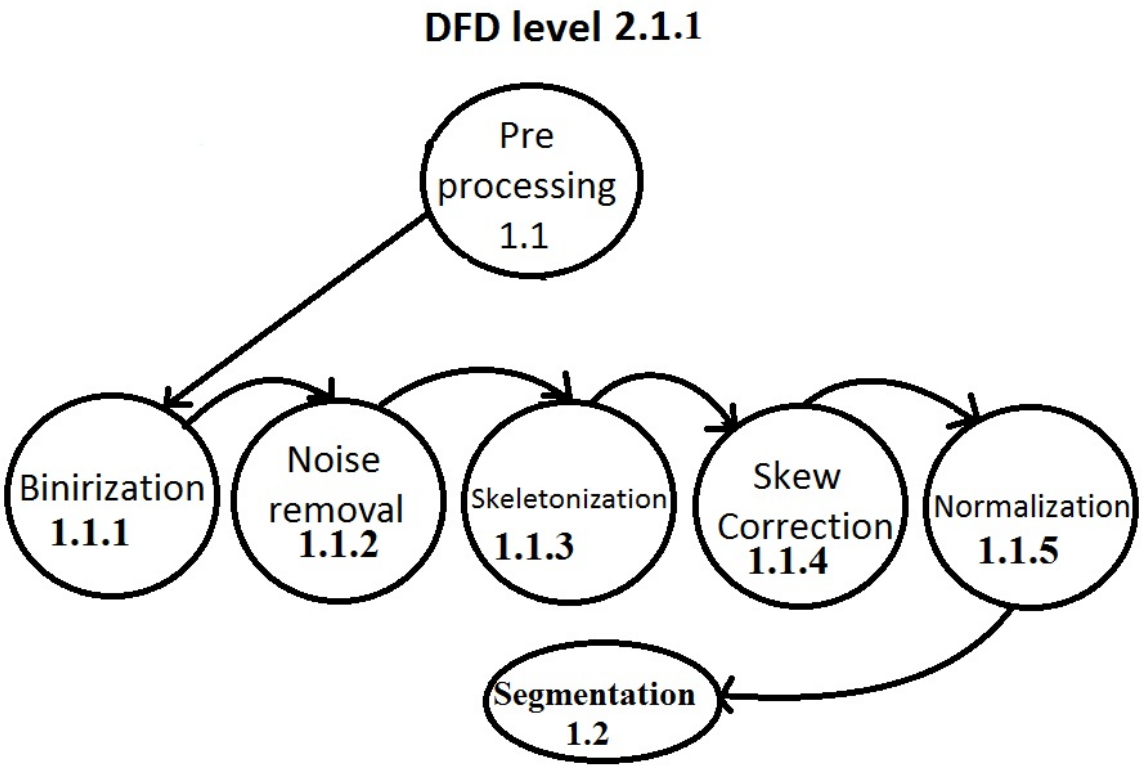



Fig. 4.3: Level 2.1.1 DFD

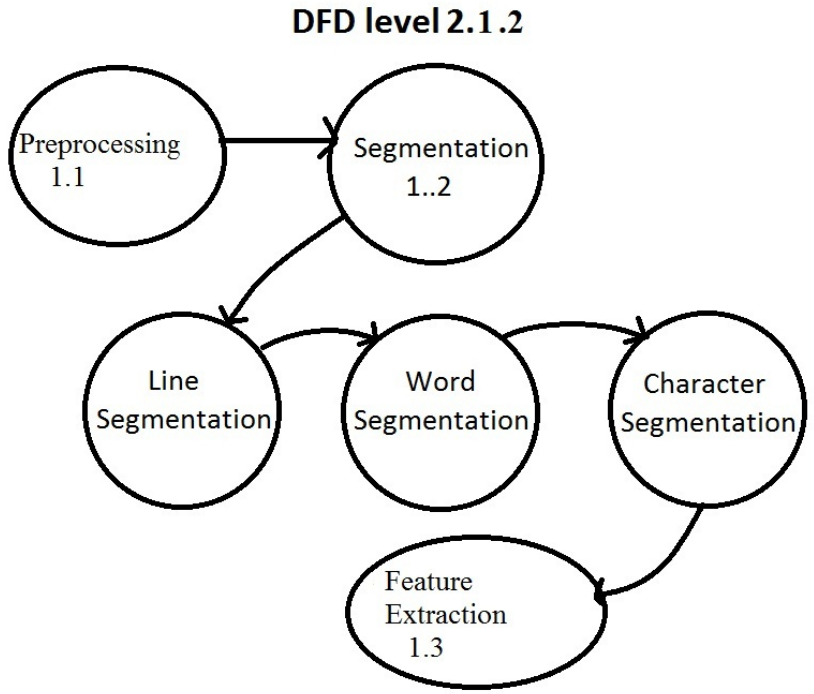



Fig. 4.4: Level 2.1.2 DFD

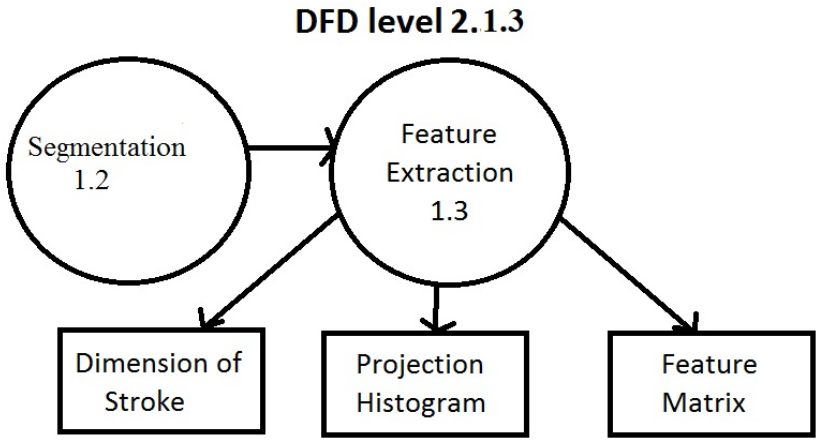



Fig. 4.5: Level 2.1.3 DFD

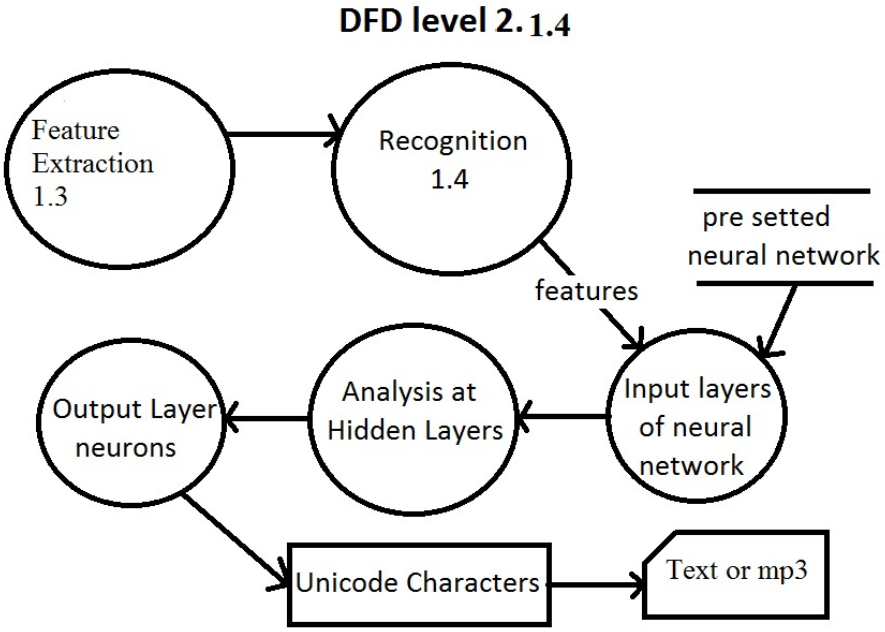



Fig. 4.6: Level 2.1.4 DFD

# Chapter 5

# METHODOLOGIES

The product is completely been developed in Java environment in Windows platform. This chapter may explain our implementation details.

## 5.1 Languages And platforms Used

The product is developing in Java as well as C++ environment. We are looking forward on Netbeans IDE 7.1 for networking and web page development. For interface we use swing and awt packages in Java8.0. The development will be carried out in LINUX platform and since we are depending only on Open Source softwares we may claim that we are technically feasible. The reason for selecting C++ as the language is been listed below. C++ is a powerful and flexible Open Source language.We use a library called **FANN**(Fast Artificial Neural Networks) to implement our neural network. Reason for using Java for developing the interface is given below. Java is developed in the name of Green project, its Write Once Read Anywhere (WORA) feature makes it a powerful language packed with cross-platform dependence. Closely linked with the Object Oriented concept but not completely Object Oriented Language due to its Interface libraries, this language offers more security than any other languages do. It is the same platform independence feature of JAVA that tempts programmers and developers to build the networking application interfaces in J2EE environment.

# Chapter 6

# CONCLUSION

Optical Character Recognition has it own significance nowadays. It has wide range of applications in the domain of postal automation, automatic number plate recognition, preservation of handwritten historical documents,bank check processing,reading aid for the blind people etc. Handwritten character recognition is traditionally divided into on-line and online recognition.We are doing an on-line HCR system.

Even though HCR systems are well advanced in foreign language scripts like Chinese and Japanese, only few works exist in Indian scripts especially in the South Indian scripts. This is mainly due to its large character set, high degree of similarity between these characters and the presence of compound characters in these scripts. Also, variations in the writing styles of different people make the recognition process more difficult. Among the Indian languages, most of the work has been reported in Devanagari and Bangla. The research on South Indian scripts namely Malayalam, Tamil, Telugu and Kannada have gained much popularity recently as many agencies like the Ministry of Communications and Information Technology and Government of India are providing aid and financial support for many of these projects.

# Chapter 7

# REFERENCES


[1] Yalniz, I.Z.; Manmatha. R, "A Fast Alignment Scheme for Automatic OCR Evaluation of Books" International Conference Document Analysis and Recognition (ICDAR), 2011.

[2] Jeff Heaton, "Introduction to neural networks in Java, Heaton Research Inc., 2005,0-9773206-0-X. [10]. Raman Maini and himanshuAggarwal, A comprehensive review of image enhancement techniques", Journal of computing, volume 2, issue 3, 2010, ISSN 2151-9617.

[3] Brijmohan Singh, Mridula, Vivek Chand, Ankush Mittal and D.Ghosh, "A comparative study of different approaches of noise removal for document images", International conference on soft computing for problem solving, volume 130/2102, p 847-854, 2011.

[4] ErginaKavallieratou and FotisDaskas,"Text line detection and segmentation Uneven skew angles and kill-and-dale writing", Journal of Universal computer science, Vol 17, p 16-29, 2010.

[5] Vatsal H. Shah"Machine Learning Techniques for Stock Prediction"

[6] Robert Burbidge, Bernard Buxton, "An Introduction to Support Vector Machines for Data Mining" *Computer Science Dept., UCL, Gower Street, WC1E 6BT, UK*

[7] C N Aganthopolous, "A License Plate Recognition algorithm for Intelligent Transportation System applications". University of the Aegean and National Technical University of Athens, p 377- 392, 2006.