

Major Project Report On

**AN AUTOMATED ANNOTATION SYSTEM FOR DOCUMENT TEXT  
IMAGES**

Submitted in partial fulfillment of the requirements for the award of the

**Bachelor of Technology**

In

**Department of Computer Science and Engineering**

By

**B. JHANSI LAKSHMI** **20241A0566**

**B. SRIJA** **20241A0564**

**MARIA JABEEN** **20241A0593**

**K. KUSHI REDDY** **20241A0583**

Under the Esteemed guidance of

**Dr. ASHLIN DEEPA R N Associate Professor**



**Department of Computer Science and Engineering**

**GOKARAJU RANGARAJU INSTITUTE OF ENGINEERING AND  
TECHNOLOGY**

**(Autonomous)**



GOKARAJU RANGARAJU  
INSTITUTE OF ENGINEERING AND TECHNOLOGY  
**(Autonomous)**

## CERTIFICATE

This is to certify that the major project entitled "**An Automated Annotation System for Document Text Images**" is submitted by **B. Jhansi Lakshmi (20241A0566), B. Srija (20241A0564), Maria Jabeen(20241A0593), K Kushi Reddy(20241A0583)** in partial fulfillment of the award of degree in BACHELOR OF TECHNOLOGY in Computer Science and Engineering during academic year 2023- 2024.

INTERNAL GUIDE	HEAD OF THE DEPARTMENT
<b>Dr. ASHLIN DEEPA R N</b>	<b>Dr. B. SANKARA BABU</b>
Associate Professor	Professor and HOD

EXTERNAL EXAMINER

## **ACKNOWLEDGEMENT**

There are many people who helped us directly and indirectly to complete our project successfully. We would like to take this opportunity to thank one and all. First we wish to express our deep gratitude towards our internal guide **Dr. Ashlin Deepa R N, Associate Professor**, Department of CSE for her/his support in the completion of our project report. We wish to express our honest and sincere thanks to **Dr. B. Sankara Babu, HOD, Department of CSE** and to our principal **Dr. J. Praveen** for providing the facilities to complete the major project. We would like to thank all our faculty and friends for their help and constructive criticism during the project completion phase. Finally, we are very much indebted to our parents for their moral support and encouragement to achieve goals.

**B. Jhansi Lakshmi (20241A0566)**

**B. Srija (20241A0564)**

**Maria Jabeen(20241A0593)**

**K Kushi Reddy(20241A0583)**

## **DECLARATION**

We hereby declare that the major project entitled "**An Automated Annotation System for Document Text Images**" is the work done during the period from **19<sup>th</sup> July 2023 to 23<sup>rd</sup> Dec 2023** and is submitted in the partial fulfillment of the requirements for the award of degree of Bachelor of Technology in Computer Science and Engineering from Gokaraju Rangaraju Institute of Engineering and Technology (Autonomous under Jawaharlal Nehru Technology University, Hyderabad).The results embodied in this project have not been submitted to any other university or Institution for the award of any degree or diploma.

**B. Jhansi Lakshmi (20241A0566)**

**B. Srija (20241A0564)**

**Maria Jabeen(20241A0593)**

**K Kushi Reddy(20241A0583)**

<b>Table of Contents</b>		
<b>Chapter</b>	<b>TITLE</b>	<b>Page No</b>
	<b>Abstract</b>	<b>1</b>
<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Literature Survey</b>	<b>3</b>
<b>3</b>	<b>Modules &amp; UML diagram</b>	<b>5</b>
<b>4</b>	<b>Implementation</b>	<b>9</b>
<b>5</b>	<b>Test cases</b>	<b>13</b>
<b>6</b>	<b>Conclusion &amp; Scope</b>	<b>19</b>
	<b>References</b>	<b>20</b>
	<b>Appendix</b> i)Review Paper ii)Result	<b>22</b>

## **Abstract**

Annotation involves enriching text with additional details or labels to provide context, improve comprehension, or aid in refining Natural Language Processing (NLP) models. The central aim of this project is to develop a robust system capable of annotating text images in documents with minimal human intervention. This involves delineating distinct regions within an image and generating descriptive text that effectively captures the characteristics of these outlined areas. While manual annotation is both time-consuming and error-prone, automatic image annotation tools step in to address these challenges, significantly boosting efficiency and reliability. A primary emphasis is placed on achieving word-level annotations for handwritten document images. The resulting system not only alleviates the time and resource demands associated with manual annotation but also elevates the reliability and accuracy of annotations. This project is poised to make a substantial contribution to the advancement of NLP applications, particularly in scenarios that involve the analysis and understanding of handwritten documents.

# **Chapter 1**

## **Introduction**

Optical Character Recognition (OCR) stands as a transformative technology crucial for converting printed or handwritten text from physical documents or images into editable and machine-readable formats. This technology plays a pivotal role in digitizing books, articles, historical records, and various textual materials. The field of OCR has witnessed significant advancements[16].

While OCR primarily focuses on printed text[15], Handwritten Text Recognition (HTR) takes text recognition a step further by specializing in deciphering handwritten content. An optimized OCR system tailored for handwritten Marathi text document classification and recognition is introduced in [17]. Handwritten materials, such as historical documents and manuscripts, pose unique challenges due to variations in writing styles, ink fading, and paper degradation. HTR technology utilizes neural networks, machine learning algorithms, and pattern recognition techniques to transcribe handwritten content, thereby preserving the authenticity and historical significance of such documents with the primary goal of converting them into a digital format[4].

Although handwritten documents provide valuable historical and cultural information, manually annotating and transcribing them is a time-consuming and error-prone task. Automated Annotation, characterized by superior efficiency and user-friendly interfaces, surpasses manual annotation. Automated Annotation Systems[5] streamline and enhance tasks related to text analysis, information retrieval, and data organization. They leverage techniques from Natural Language Processing (NLP), machine learning, and data mining to analyze and understand text, facilitating the automated generation of tags for large volumes of textual content.

An innovative automatic annotation-based approach for digitized text recognition is proposed in [10]. Manual annotation is labor-intensive, time-consuming, and error-prone. Automated annotation systems provide an efficient and user-friendly alternative, expediting the annotation process while improving accuracy. These systems find diverse applications, including content recommendation, information retrieval, sentiment analysis, and content categorization.

## **Chapter 2**

### **Literature Survey**

In this digital age, the need for efficient document digitization and text extraction has led to the development of Optical Character Recognition (OCR) systems. These systems aim to automatically convert printed or handwritten text from document images into machine-readable and searchable text. One crucial aspect of enhancing OCR accuracy and usability is the automatic annotation of document text images.

Choi and Kim (2012) delve into automatic image annotation using semantic text analysis[3]. Their work highlights the intersection of image analysis and natural language processing, demonstrating how semantic understanding of text can aid in annotating images.

Hu et al. (2015) introduce DocRicher[2] an automatic annotation system that leverages social media to annotate text documents. This innovative approach showcases the significance of incorporating external contextual information, such as social media trends or user-generated content, to annotate and enrich document text.

In the study by[1] Mondal et al. (2023), the authors focus on handwritten document images and propose a method for automatic annotation at the word level. Automatic annotation systems play a pivotal role in enhancing the accuracy and usability of OCR technology. They contribute to the enrichment of document text images with metadata, semantic context, and linguistic information.

[6,7] uses a hybrid CNN and RNN architecture that tackles the challenges of recognizing handwritten and printed text.

In the realm of Optical Character Recognition (OCR), recognizing handwritten paragraphs poses a significant challenge. Language models like BERT and GPT for Indian languages, offering insights into their adaptability and performance in multilingual NLP tasks. [8] addresses this challenge by introducing an innovative approach using a Vertical Attention Network (VAN). VAN employs an attention mechanism focused on vertical text regions, enabling it to capture the intricate layout and structure of handwritten paragraphs seamlessly, thereby eliminating the need for explicit segmentation.

Traditional metrics like WER and CER are widely used in OCR experiments for assessing recognition accuracy[13].The paper introduces the concept of bWER[14], a metric designed to assess word recognition independently of reading order.

Dealing with limited labeled data, especially for symbols or alphabets in regional languages, presents a common challenge in HTR. [9] introduces a few-shot learning approach for HTR, particularly beneficial when data availability is scarce.This approach starts with a minimal number of labeled examples for each symbol and progressively refines its understanding as more labeled data becomes accessible. To overcome the data limitation, the paper introduces the technique of Pseudo-labeling.

Automatic annotation systems have direct applications in OCR technology. By associating meaningful annotations with text images, these systems improve various aspects of OCR, including text recognition accuracy, context preservation tasks.

# **Chapter 3**

## **Modules and UML diagram**

### **Modules:**

There are three main modules in our project.

#### **1) De-GAN Module:**

- De-GAN utilizes Generative Adversarial Networks (GANs) for denoising handwritten documents.
- It identifies and removes various types of noise, such as background artifacts and smudges.
- It enhances the overall quality of scanned documents, making them more legible and suitable for annotation.
- It employs deep learning techniques to learn and generate denoised versions of input documents.
- It significantly improves dataset readiness for subsequent annotation and analysis.

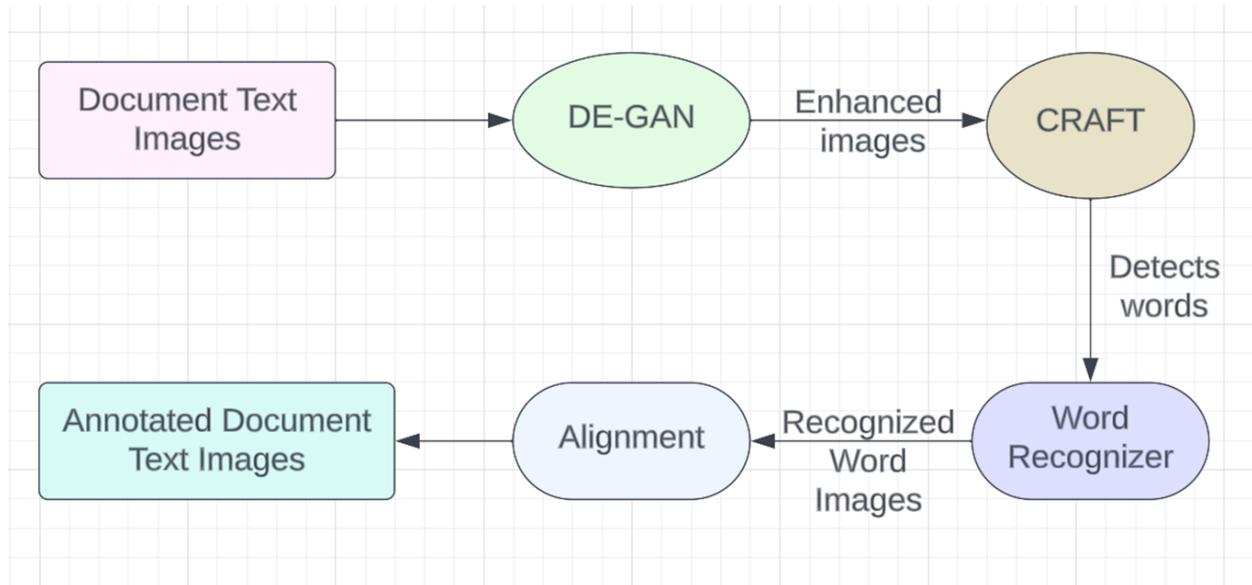
#### **2) CRAFT Module:**

- After binarizing the image using De-GAN, it is sent to the CRAFT model.
- CRAFT is a text detection model that uses a Convolutional Neural Network(CNN) model to accurately extract text regions from the document images.
- It analyzes pixel-level information to determine character and word boundaries.
- It utilizes computer vision algorithms to identify regions containing textual content.
- The purpose of this model is to minimize errors and enhance annotation efficiency by focusing solely on relevant text.

### 3) Annotation Module:

- Annotation involves the automated labeling of pre-processed documents to make it recognizable and usable for various applications.
- It employs machine learning models like CNNs, GCN, LSTM-based and encoder-decoder with attention models for automated text recognition.
- It involves manual review to rectify discrepancies and ensure high accuracy.

### Architecture:



**Fig 1: Architecture of Automated Annotation System**

## UML Diagrams:

### Use Case Diagram:

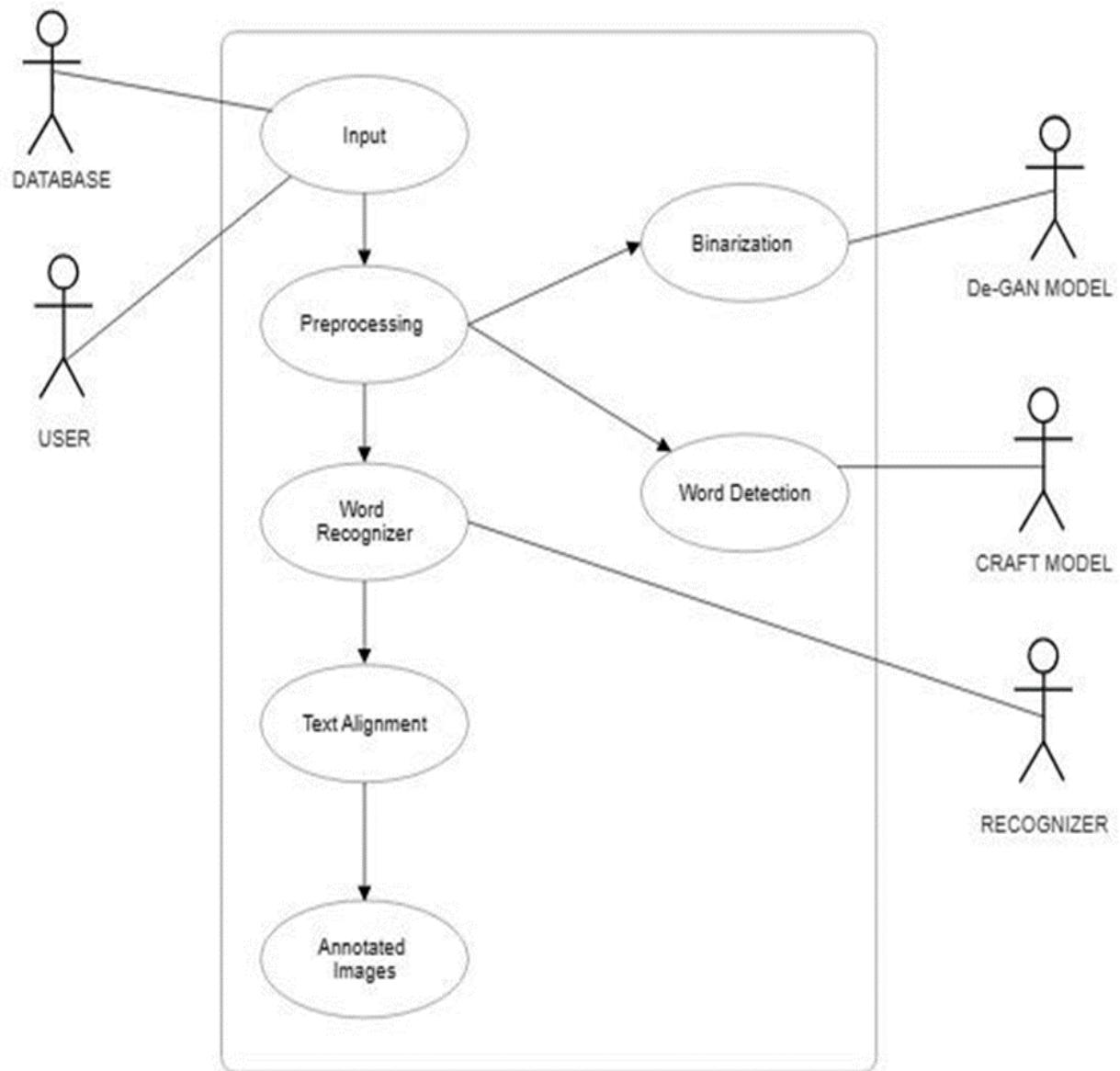


Fig 2: Use Case Diagram of Automated Annotation System

### Data Flow Diagram:

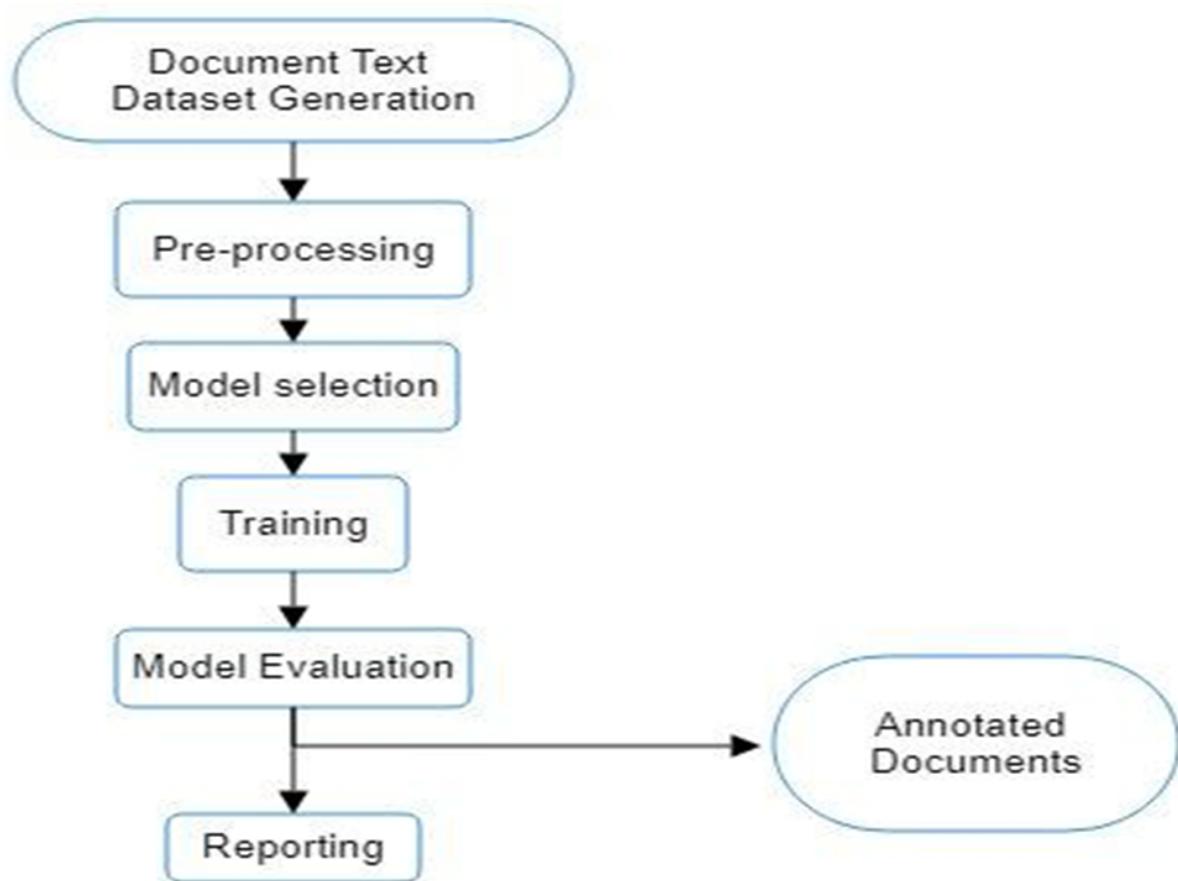


Fig 3: Data Flow Diagram of Automated Annotation System

# Chapter 4

## Implementation

### Implementation steps:

- The process begins with the acquisition of an image, ensuring it is taken without the influence of any filters.
  - The unfiltered image is fed into the pretrained De-GAN (Denoising GAN) model.
  - De-GAN specializes in denoising and enhances the image quality, generating a binarized version that eliminates background noises.
  - The binarized image produced from De-GAN serves as input for the CRAFT model, which generates a bounding box image and a mask image highlighting detected text regions.
- Additionally, a text file is created, containing coordinates detailing the position of bounding boxes.

### De-GAN snapshots :

```
1 2 #!/usr/bin/env python
3 import sys
4 import numpy as np
5 import matplotlib.pyplot as plt
6 import scipy.misc
7 import math
8 from PIL import Image
9 import random
10 from utils import *
11 from models import *
12
13 input_size = (256,256,1)
14
15 task = sys.argv[1]
16
17 codes_folder_path = 'drive/MyDrive/Codes/'
18 if task =='binarize':
19     generator = generator_model(biggest_layer=1024)
20     generator.load_weights(codes_folder_path+"binarization_generator_weights.h5")
21
22
23 deg_image_path = sys.argv[2]
24
25 deg_image = Image.open(deg_image_path)# /255.0
26 deg_image = deg_image.convert('L')
27 deg_image.save('curr_image.png')
28
29
```

**Fig 4**

```
29
30 test_image = plt.imread('curr_image.png')
31
32
33
34
35 h = ((test_image.shape [0] // 256) +1)*256
36 w = ((test_image.shape [1] // 256 ) +1)*256
37
38 test_padding=np.zeros((h,w))+1
39 test_padding[:test_image.shape[0],:test_image.shape[1]]=test_image
40
41 test_image_p=split2(test_padding.reshape(1,h,w,1),1,h,w)
42 predicted_list=[]
43 for l in range(test_image_p.shape[0]):
44     | predicted_list.append(generator.predict(test_image_p[l].reshape(1,256,256,1)))
45
46 predicted_image = np.array(predicted_list).reshape()
47 predicted_image=merge_image2(predicted_image,h,w)
48
49 predicted_image=predicted_image[:test_image.shape[0],:test_image.shape[1]]
50 predicted_image=predicted_image.reshape(predicted_image.shape[0],predicted_image.shape[1])
51
52
53 if task == 'binarize':
54     | bin_thresh = 0.95
55     | predicted_image = (predicted_image[:, :, :]>bin_thresh)*1
56
57
58 save_path = sys.argv[3]
59 plt.imsave(save_path, predicted_image,cmap='gray')
60
61
```

**Fig 5**

## Craft code snapshots:

```
# -*- coding: utf-8 -*-
import torch
import torch.nn as nn
import torch.nn.functional as F
from basenet.vgg16_bn import vgg16_bn, init_weights

class double_conv(nn.Module):
    def __init__(self, in_ch, mid_ch, out_ch):
        super(double_conv, self).__init__()
        self.conv = nn.Sequential(
            nn.Conv2d(in_ch + mid_ch, mid_ch, kernel_size=1),
            nn.BatchNorm2d(mid_ch),
            nn.ReLU(inplace=True),
            nn.Conv2d(mid_ch, out_ch, kernel_size=3, padding=1),
            nn.BatchNorm2d(out_ch),
            nn.ReLU(inplace=True)
        )

    def forward(self, x):
        x = self.conv(x)
        return x
```

Fig 6

```
if __name__ == '__main__':
    model = CRAFT(pretrained=True).cuda()
    output, _ = model(torch.randn(1, 3, 768, 768).cuda())
    print(output.shape)
```

Fig 7

```

class CRAFT(nn.Module):
    def __init__(self, pretrained=False, freeze=False):
        super(CRAFT, self).__init__()

        """ Base network """
        self.basenet = vgg16_bn(pretrained, freeze)

        """ U network """
        self.upconv1 = double_conv(1024, 512, 256)
        self.upconv2 = double_conv(512, 256, 128)
        self.upconv3 = double_conv(256, 128, 64)
        self.upconv4 = double_conv(128, 64, 32)

        num_class = 2
        self.conv_cls = nn.Sequential(
            nn.Conv2d(32, 32, kernel_size=3, padding=1), nn.ReLU(inplace=True),
            nn.Conv2d(32, 32, kernel_size=3, padding=1), nn.ReLU(inplace=True),
            nn.Conv2d(32, 16, kernel_size=3, padding=1), nn.ReLU(inplace=True),
            nn.Conv2d(16, 16, kernel_size=1), nn.ReLU(inplace=True),
            nn.Conv2d(16, num_class, kernel_size=1),
        )

        init_weights(self.upconv1.modules())
        init_weights(self.upconv2.modules())
        init_weights(self.upconv3.modules())
        init_weights(self.upconv4.modules())
        init_weights(self.conv_cls.modules())

    def forward(self, x):
        """ Base network """
        sources = self.basenet(x)

        """ U network """
        y = torch.cat([sources[0], sources[1]], dim=1)
        y = self.upconv1(y)

        y = F.interpolate(y, size=sources[2].size()[2:], mode='bilinear', align_corners=False)
        y = torch.cat([y, sources[2]], dim=1)
        y = self.upconv2(y)

        y = F.interpolate(y, size=sources[3].size()[2:], mode='bilinear', align_corners=False)
        y = torch.cat([y, sources[3]], dim=1)
        y = self.upconv3(y)

        y = F.interpolate(y, size=sources[4].size()[2:], mode='bilinear', align_corners=False)
        y = torch.cat([y, sources[4]], dim=1)
        feature = self.upconv4(y)

        y = self.conv_cls(feature)

        return y.permute(0,2,3,1), feature

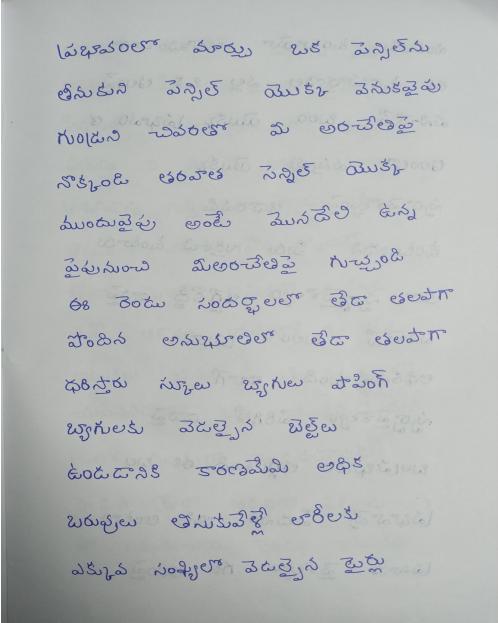
```

**Fig 8**

## Chapter 5

### Test cases

#### De-GAN Module:

ID	Test Data	Output	Status(pass /Fail)
1	 <p>ಮಾರ್ಕೆಟಿಂಗ್ ಮಾರ್ಕೆಟ್ ಬಕ ಪನ್ಯೋಗಳನ್ನು ಶೈಲಿಗಳನ್ನು ಪನ್ಯೋಗ್ ಯೊಕ್ಕ ವಸ್ತುಗಳನ್ನು ಸಂಪೂರ್ಣ ರೀವಿಯ್ಸ್ ಮತ್ತು ಅರ್ಥಾತ್ ಅಂತರ್ರಂಭ ನೊಟ್‌ಎಡಿಟ್ ಅಂದಿನ ಪನ್ಯೋಗ್ ಯೊಕ್ಕ ವಸ್ತುಗಳನ್ನು ಮುಂದುವರ್ತು ಅಂದೆ ಮೊನದೆಲ್ ಟಂಸ್ ಪ್ರೈಸ್‌ನಿಂದ ಮೀಲಿಂಗ್‌ನಿಂದ ಸುಧ್ಯಾರ್ಥ ಇಂ ರೆಂಡ್ ಸಂದರ್ಭದಲ್ಲಾ ತೆದೆ ಉಂಪಾಗ್ ಪಂಹಿನ ಅನುಭೂತಿಗ್ ತೆದೆ ಉಲಂಘಾ ಧರಿಸ್ತರು ಸುಜ್ಯಾಲ್ ಚ್ಯಾಗ್ಸ್ ಹಾಸ್ಟಿಂಗ್ ಚ್ಯಾಗ್ಸ್‌ಲಕ್ಟ್ ವೆಡಲ್ಟ್‌ನ ಬೆಲ್ಟ್‌ಲ ಟಂಡಂಾಸಿನ್ ಕಾರ್ಬಾಮೆನ್ ಅಥಿಕ್ ಬರ್ಲಿನ್ ಶೈಲಿಗಳನ್ನು ಲೋರ್‌ಲಿ ಎಕ್ಸ್‌ಪ್ರೈಸ್ ಸಂಕ್ಷಿಪ್ತ್ ವೆಡಲ್ಟ್‌ನ ತ್ವರ್ತಿ</p>	ಮಾರ್ಕೆಟಿಂಗ್ ಮಾರ್ಕೆಟ್ ಬಕ ಪನ್ಯೋಗಳನ್ನು ಶೈಲಿಗಳನ್ನು ಪನ್ಯೋಗ್ ಯೊಕ್ಕ ವಸ್ತುಗಳನ್ನು ಸಂಪೂರ್ಣ ರೀವಿಯ್ಸ್ ಮತ್ತು ಅರ್ಥಾತ್ ಅಂತರ್ರಂಭ ನೊಟ್‌ಎಡಿಟ್ ಅಂದಿನ ಪನ್ಯೋಗ್ ಯೊಕ್ಕ ವಸ್ತುಗಳನ್ನು ಮುಂದುವರ್ತು ಅಂದೆ ಮೊನದೆಲ್ ಟಂಸ್ ಪ್ರೈಸ್‌ನಿಂದ ಮೀಲಿಂಗ್‌ನಿಂದ ಸುಧ್ಯಾರ್ಥ ಇಂ ರೆಂಡ್ ಸಂದರ್ಭದಲ್ಲಾ ತೆದೆ ಉಂಪಾಗ್ ಪಂಹಿನ ಅನುಭೂತಿಗ್ ತೆದೆ ಉಲಂಘಾ ಧರಿಸ್ತರು ಸುಜ್ಯಾಲ್ ಚ್ಯಾಗ್ಸ್ ಹಾಸ್ಟಿಂಗ್ ಚ್ಯಾಗ್ಸ್‌ಲಕ್ಟ್ ವೆಡಲ್ಟ್‌ನ ಬೆಲ್ಟ್‌ಲ ಟಂಡಂಾಸಿನ್ ಕಾರ್ಬಾಮೆನ್ ಅಥಿಕ್ ಬರ್ಲಿನ್ ಶೈಲಿಗಳನ್ನು ಲೋರ್‌ಲಿ ಎಕ್ಸ್‌ಪ್ರೈಸ್ ಸಂಕ್ಷಿಪ್ತ್ ವೆಡಲ್ಟ್‌ನ ತ್ವರ್ತಿ	Pass

2	<p>ఎందుకు పుంచూమూ శిల్పాచించారా ..</p> <p>ఈ ట్రియాపాల వల్ల ఒక తపంపై పనిచేసే బిలం యొక్క ప్రభావం ఆ తపంటో మీవు యొక్క స్ట్రోక్షెల్చుపై శిఫరవది ఉండుందని మీరు సత్కాచి పుంచారు కదా స్ట్రోక్షెల్చుపై శిఫరవది దానపై పనిచేసే నాబిలం అప్పించాలవం అధికములుండి ఆలగీ స్ట్రోక్షెల్చుపై వెరిటీ దానికి బిలపు భావం తెగ్గించుంది ఈ బిల ప్రభావాన్ని లపించును అని అంచం క్రమాను తోల్చుంచు గల తపంపై</p>	<p>ఎందుకు పుంచూమూ శిల్పాచించారా ..</p> <p>ఈ ట్రియాపాల వల్ల ఒక తపంపై పనిచేసే బిలం యొక్క ప్రభావం ఆ తపంటో మీవు యొక్క స్ట్రోక్షెల్చుపై శిఫరవది ఉండుందని మీరు సత్కాచి పుంచారు కదా స్ట్రోక్షెల్చుపై శిఫరవది దానపై పనిచేసే బిలం అప్పించాలవం అధికములుండి ఆలగీ స్ట్రోక్షెల్చుపై వెరిటీ దానికి బిలపు భావం తెగ్గించుంది ఈ బిల ప్రభావాన్ని పెదునం అని అంచం క్రమాను “తోల్చుంచు గల తపంపై</p>	Pass
3	<p>అది తన గల్వొన దినను మాస్ట్రోక్షెల్చుంచు వెరిటీలీంచండి వటం వటు వ వటుం ఎ నెచ్చులప్పితెలు ఉన్న బంతిష్ట బలాన్ని పెయ్యోగించడం వటుం చి కములు ఉన్న బంతిష్ట అదే దినలు మంకల బలాన్ని పెయ్యోగించడం క్రి శి నెచ్చులప్పితెలు ఉన్న బంతిష్ట బలాన్ని పెయ్యోగించడం వ్వోరా దొనిని కదపవచ్చు అలాచే కములుతున్న బంతిని వెప్పుకొవడం వ్వోరా నెచ్చులప్పితే తీసుకురాగులు అని మీరు గమనించి ఉంచారు ఈ</p>	<p>అది తన గల్వొన దినను మాస్ట్రోక్షెల్చుంచు వెరిటీలీంచండి వటం వటు వ వటుం ఎ నెచ్చులప్పితెలు ఉన్న బంతిష్ట బలాన్ని పెయ్యోగించడం వటుం చి కములు ఉన్న బంతిష్ట అదే దినలు మంకల బలాన్ని పెయ్యోగించడం క్రి శి నెచ్చులప్పితెలు ఉన్న బంతిష్ట బలాన్ని పెయ్యోగించడం వ్వోరా దొనిని కదపవచ్చు అలాచే కములుతున్న బంతిని వెప్పుకొవడం వ్వోరా నెచ్చులప్పితే తీసుకురాగులు అని మీరు గమనించి ఉంచారు ఈ</p>	Fail

4	<p>బంబగ పనచేసే బలాన్న వీడును అంటం వీడును బలం షైల్యం వీడును దుక్క శ్రవణాం స్వాచున్ మీటు ఇ మి క్ కరం బలం శ్రభావాన్ గృత్తింయుట రెండు శ్రీలు శీసుకుని వాటిని హోదిసున్నంతో తేయ మెత్తని ఇసుకుతో నింపంది ఒకే తెకరం దెవ్యరాశి గల రెండు ఇటుకులు తెసుక్కండి భి పటుం లో చుమ్మినిధంగా ఒక జుటుకుని మొదటి ప్రైల్ నిఱులుగా రెండువదానని రెండవ ప్రైల్ అష్టంగా పెట్టుంది ఏ గమనించారు రెండు జుటుకులు</p>	<p>బంబగ పనచేసే బలాన్న వీడును అంటం వీడును బలం షైల్యం వీడును దుక్క శ్రవణాం స్వాచున్ మీటు ఇ మి క్ కరం బలం శ్రభావాన్ గృత్తింయుట రెండు శ్రీలు శీసుకుని వాటిని హోదిసున్నంతో తేయ మెత్తని ఇసుకుతో నింపంది ఒకే తెకరం దెవ్యరాశి గల రెండు ఇటుకులు తెసుక్కండి భి పటుం లో చుమ్మినిధంగా ఒక జుటుకుని మొదటి ప్రైల్ నిఱులుగా రెండువదానని రెండవ ప్రైల్ అష్టంగా పెట్టుంది ఏ గమనించారు రెండు జుటుకులు</p>	Pass
5	<p>సుస్థంలోకి ఒకేయ లోటుకి డిగాయి తేయ ఎలుదుకు పటుం ప్రైల్ నిఱులుగా ఉంచిన జుటుక తుఫ్టంగా ఉంచిన జుటుక కంటే హోదిసున్నం ఎక్కువ లోటుగా డిగిందని గమనిస్తం రెండు జుటుకులూ ఒకే బయలుని కలిగి ఉన్నయి కాబట్టి రెండు ప్రైల్లా గల సుస్థంపై లవి కలగచేసే బలం సమానంగా ఉంటుంది కాన సున్నాన్ని తో జుటుకుల ఉపరితల షైల్యాల్లు మార్పి లవి ఎంచ లోటు దిగెళ్ళాయా నీర్మయస్తుంది తై</p>	<p>సుస్థంలోకి ఒకే లోటుకి డిగాయి తేయ ఎలుదుకు పటుం ప్రైల్ నిఱులుగా ఉంచిన జుటుక తుఫ్టంగా ఉంచిన జుటుక కంటే హోదిసున్నం ఎక్కువ లోటుగా డిగిందని గమనిస్తం రెండు జుటుకులూ ఒకే బయలుని కలిగి ఉన్నయి కాబట్టి రెండు ప్రైల్లా గల సుస్థంపై లవి కలగచేసే బలం సమానంగా ఉంటుంది కాన సున్నాన్ని తో జుటుకుల ఉపరితల షైల్యాల్లు మార్పి లవి ఎంచ లోటు దిగెళ్ళాయా నీర్మయస్తుంది తై</p>	Pass

## CRAFT Module:

ID	Test Data	Output	Status(pass /Fail)
1	<p>గ్రాఫికంలో మూడు ఒక పెస్టిల్సు లీసుకుని పెస్సిల్ యొక్క పెనుక్కుతుపు గుండులు చెంది మ అరచేతిట్లు నొక్కండి తెరవాళ సెస్సిల్ యొక్క ముందువైపు అంటే మునదేలి ఉన్న ప్రత్యుషమయి మీలరచేతిట్లు గుండులీ ఈ రెండు సందర్భాలలో తేడా తమిగా ప్రంహిన అనుభూతిలో తేడా తెలపాగా ధరిస్తామ సుశుల చ్యాగుల పోంగి చ్యాగులకు వెడత్తేన చెట్టుల ఉండడానికి కారణమేమి అధిక బమ్బులు లైసెస్చెచ్చి లరీలకు ఎండ్లు సంఖ్యలో వెడత్తేన త్లులు ..</p>	<p>గ్రాఫికంలో మూడు ఒక పెస్టిల్సు లీసుకుని పెస్సిల్ యొక్క పెనుక్కుతుపు గుండులు చెంది మ అరచేతిట్లు నొక్కండి తెరవాళ సెస్సిల్ యొక్క ముందువైపు అంటే మునదేలి ఉన్న ప్రత్యుషమయి మీలరచేతిట్లు గుండులీ ఈ రెండు సందర్భాలలో తేడా తమిగా ప్రంహిన అనుభూతిలో తేడా తెలపాగా ధరిస్తామ సుశుల చ్యాగుల పోంగి చ్యాగులకు వెడత్తేన చెట్టుల ఉండడానికి కారణమేమి అధిక బమ్బులు లైసెస్చెచ్చి లరీలకు ఎండ్లు సంఖ్యలో వెడత్తేన త్లులు ..</p>	Pass

2	<p>ఎందుకు , ఉండూయో తెల్చించారు . . .</p> <p>ఈ ఉదహరణల పట్ల ఒక ఉపాంశు</p> <p>పనిచేసే బఱం ; మొక్క , శుభామం ఈ</p> <p>ఉంటో .. యస్తువు యొక్క</p> <p>స్వాధైకౌల్యంపై ఆధారవది</p> <p>ఉంటుందని , మీరు గుర్తించి ఉంటారు</p> <p>కదా స్వాధైకౌల్యం . తగ్గికైంది దానట</p> <p>పనిచేసే .. బఱం .. శుభామం</p> <p>అధికమతుంది.. , అలగో ..</p> <p>స్వాధైకౌల్యం .. వెరిసెటె దానట</p> <p>బఱపథం తగ్గిమాటి కాకి బఱ ..</p> <p>త్రుభావాన్ని పెదునం .. అని అంటాం ..</p> <p>త్రుమాన .. త్రైకౌల్యం .. గల ఉపాంశు .</p>	<p>ఎందుకు , ఉండూయో తెల్చించారు . . .</p> <p>ఈ ఉదహరణల పట్ల ఒక ఉపాంశు</p> <p>పనిచేసే బఱం ; మొక్క , శుభామం ఈ</p> <p>ఉంటో .. యస్తువు యొక్క</p> <p>స్వాధైకౌల్యంపై ఆధారవది</p> <p>ఉంటుందని , మీరు గుర్తించి ఉంటారు</p> <p>కదా స్వాధైకౌల్యం . తగ్గికైంది దానట</p> <p>పనిచేసే .. బఱం .. శుభామం</p> <p>అధికమతుంది.. , అలగో ..</p> <p>స్వాధైకౌల్యం .. వెరిసెటె దానట</p> <p>బఱపథం తగ్గిమాటి కాకి బఱ ..</p> <p>త్రుభావాన్ని పెదునం .. అని అంటాం ..</p> <p>త్రుమాన .. త్రైకౌల్యం .. గల ఉపాంశు .</p>	Pass
3	<p>ఉపాంశు తన నవ్వున దీనను</p> <p>మూస్తుకుండ వెతిన్నాలించండి.. పటువ</p> <p>వటు వ వెంటం న నెళ్లుబస్తుతిలో</p> <p>టీవ్సు బంతిష్ట బలాన్ని</p> <p>పెయ్యోగించడం వటుం చి</p> <p>కములు ల్యుస్ బంతిష్ట అడో దీనలూ</p> <p>మంకల బలాన్ని పెయ్యోగించడం త్రై</p> <p>ఈ నెళ్లుబస్తుతిలో ఈన్స్ బంతిష్టే</p> <p>బలాన్ని పెయ్యోగించడం వ్యార్లా</p> <p>చూసిని కమవపచ్చ అలాడే</p> <p>కములుతున్స్ బంతిష్టి పట్టుకొంచడం</p> <p>ట్యూరా నెళ్లుబస్తుతిఁ త్రేసుకుంచటం</p> <p>అని మీరు గమనించి ఉంటారు త్రు</p>	<p>ఉపాంశు తన నవ్వున దీనను</p> <p>మూస్తుకుండ వెతిన్నాలించండి.. పటువ</p> <p>వటు వ వెంటం న నెళ్లుబస్తుతిలో</p> <p>టీవ్సు బంతిష్ట బలాన్ని</p> <p>పెయ్యోగించడం వటుం చి</p> <p>కములు ల్యువ్ బంతిష్ట ఇద్ద దీనలూ</p> <p>మంకల బలాన్ని పెయ్యోగించడం త్రై</p> <p>ఈ నెళ్లుబస్తుతిలో ఈన్స్ బంతిష్టే</p> <p>బలాన్ని పెయ్యోగించడం వ్యార్లా</p> <p>చూసిని కమవపచ్చ అలాడే</p> <p>కములుతున్స్ బంతిష్టి పట్టుకొంచడం</p> <p>ట్యూరా నెళ్లుబస్తుతిఁ త్రేసుకుంచటం</p> <p>అని మీరు గమనించి ఉంటారు త్రు</p>	Fail

4	<p>అంబంగా పనిచేసే బలాన్న వీడునం అంటం పెదునం బలం త్రైప్యం పెదునం యెక్కు క్రమాశం న్యూటన్ వీటరు ఇ మి కి కరమం బలం త్రాభోష్ణ గృహించుట రెండు ప్రైలు తీసుకుని వాటిని పొడిసున్నంతో తేదా ముత్తని ఇసుటో సింపండి ఒకే త్రారం ద్రవ్యరాశి గల రెండు ఇటుకలు తొసుక్కండి ఏ పటుం లో చుమ్మినపథంగా ఒక ఇటుకని మొదటి పైలూ నియత్రగా రెండువదానని రెండు పైలూ అధ్యంగా వెళ్లండి మా గమనించారు రెండు ఇటుకలు</p>	<p>అంబంగా పనిచేసే బలాన్న వీడునం అంటం పెదునం బలం త్రైప్యం పెదునం యెక్కు క్రమాశం న్యూటన్ వీటరు ఇ మి కి కరమం బలం త్రాభోష్ణ గృహించుట రెండు ప్రైలు తీసుకుని వాటిని పొడిసున్నంతో తేదా ముత్తని ఇసుటో సింపండి ఒకే త్రారం ద్రవ్యరాశి గల రెండు ఇటుకలు తొసుక్కండి ఏ పటుం లో చుమ్మినపథంగా ఒక ఇటుకని మొదటి పైలూ నియత్రగా రెండువదానని రెండు పైలూ అధ్యంగా వెళ్లండి మా గమనించారు రెండు ఇటుకలు</p>	Pass
5	<p>సుస్సుంలోకి ఒకే లోటుక తొఱుం లేదా ఎందుకు పటుం పైలూ నియత్రగా ఉంచిన ఇటుక త్రాధ్యంగా ఉంచిన ఇటుక కుటె పొడిసున్నంలు ఎక్కువ లోటుగా తగిందని గమనిస్తాం రెండు ఇటుకలూ ఒకే బురువుని కలిగి ఉస్కుయి కాబట్టి రెండు పైలూ గల సుస్సుంటే అవి కుంపిచేసే బలం సమానంగా ఉండుండి కొని సుస్సుంలు తోడే ఇటుకలు ఉపరితల త్రైప్యాల్ఫ్రోన్ మార్పు అవి ఎంత లోటుతు దిగెళ్ళాయి నీర్దయస్తుండి తై</p>	<p>సుస్సుంలోకి ఒకే లోటుక తొఱుం లేదా ఎందుకు పటుం పైలూ నియత్రగా ఉంచిన ఇటుక త్రాధ్యంగా ఉంచిన ఇటుక కుటె పొడిసున్నంలు ఎక్కువ లోటుగా తగిందని గమనిస్తాం రెండు ఇటుకలూ ఒకే బురువుని కలిగి ఉస్కుయి కాబట్టి రెండు పైలూ గల సుస్సుంటే అవి కుంపిచేసే బలం సమానంగా ఉండుండి కొని సుస్సుంలు తోడే ఇటుకలు ఉపరితల త్రైప్యాల్ఫ్రోన్ మార్పు అవి ఎంత లోటుతు దిగెళ్ళాయి నీర్దయస్తుండి తై</p>	Pass

## **Chapter 6**

### **Conclusion and scope**

In conclusion, an Automated Annotation System designed for document text images brings transformative benefits to various industries. By leveraging advanced image processing techniques, such as Optical Character Recognition (OCR) and deep learning algorithms, this system significantly enhances document analysis, data extraction, and information retrieval processes. It optimizes efficiency, accuracy, and productivity by automating the time-consuming task of manual annotation. This system finds applications in areas ranging from digitizing historical archives and cataloging large document databases to aiding in content indexing and semantic understanding. As technology continues to advance, the development and adoption of Automated Annotation Systems for document text images hold the promise of revolutionizing the way we interact with and extract knowledge from textual documents.

## References

1. Ajoy Mondal, Krishna Tulsyan, and C.V. Jawahar. "Automatic Annotation of Handwritten Document Images at Word Level." ICVGIP '22, ACM, 2023.
2. Qiang Hu, Qi Liu, Xiaoli Wang, Anthony K.H. Tung, Shubham Goyal, and Jisong Yang. "DocRicher: An Automatic Annotation System for Text Documents Using Social Media." SIGMOD '15, ACM, 2015.
3. Choi, D., Kim, P. "Automatic Image Annotation Using Semantic Text Analysis." CD-ARES 2012, Springer, 2012.
4. Balci, Batuhan, Dan Saadati, and Dan Shiferaw. "Handwritten text recognition using deep learning." CS231n: Convolutional Neural Networks for Visual Recognition, Stanford University, Course Project Report, Spring (2017): 752-759.
5. Mondal, Ajoy & Tulsyan, Krishna & Jawahar, C.. (2022). Automatic Annotation of Handwritten Documents in Word Level. 10.1145/3571600.3571645.
6. Jain, Mohit & Mathew, Minesh & Jawahar, C.V.. (2017). Unconstrained OCR for Urdu Using Deep CNN-RNN Hybrid Networks. 747-752. 10.1109/ACPR.2017.5.
7. Gunna, Sanjana, Rohit Saluja, and C. V. Jawahar. "Transfer learning for scene text recognition in Indian languages." International Conference on Document Analysis and Recognition. Cham: Springer International Publishing, 2021.
8. Coquenet, Denis & Chatelain, Clement & Paquet, Thierry. (2020). End-to-end Handwritten Paragraph Text Recognition Using a Vertical Attention Network.

9. Souibgui, M. A., Fornés, A., Kessentini, Y., and Megyesi, B., “Few shots are all you need: A progressive learning approach for low resource handwritten text recognition”, *< i>Pattern Recognition Letters*, vol. 160, pp. 43–49, 2022. doi:10.1016/j.patrec.2022.06.003.
10. Sankar, K. Pramod and C. V. Jawahar. “Enabling Search over Large Collections of Telugu Document Images - An Automatic Annotation Based Approach.” Indian Conference on Computer Vision, Graphics & Image Processing (2006).
13. Bansal, Siddhant, Praveen Krishnan, and C. V. Jawahar. "Fused Text Recogniser and Deep Embeddings Improve Word Recognition and Retrieval." International Workshop on Document Analysis Systems. Cham: Springer International Publishing, 2020.
14. Vidal, Enrique and Toselli, Alejandro Héctor and Rios-Vila, Antonio and Calvo-Zaragoza, Jorge, End-to-End Page-Level Assessment Of Handwritten Text Recognition. Available at SSRN: <https://ssrn.com/abstract=4347108> or <http://dx.doi.org/10.2139/ssrn.4347108>
15. B M, Sagar & Shobha, G. & P., Ramakanth. (2008). OCR for printed Kannada text to machine editable format using database approach. WSEAS Transactions on Computers. 7. 766-769.
16. Sharma, Parikshit. (2023). Advancements in OCR: A Deep Learning Algorithm for Enhanced Text Recognition. International Journal of Inventive Engineering and Sciences. 10. 1-7. 10.35940/ijies.F4263.0810823.
17. Ramteke, Surendra & Gurjar, Ajay & deshmukh, Dhiraj. (2018). A streamlined OCR system for handwritten MARATHI text document classification and recognition using SVM-ACS algorithm. International Journal of Intelligent Engineering and Systems. 11. 186-195. 10.22266/ijies2018.0630.20.

# Appendix

## i)Review Paper:

### A review on Automated Annotation System for Document Text Images

#### **Abstract:**

Automated Annotation Systems for text are designed to automatically add descriptive metadata or labels to textual data. In recent years, automated annotation systems for text images, mainly handwritten text images, have gained significant attention from researchers and practitioners. This review thoroughly assesses the latest advancements in Automated Annotation Systems designed for text images. It explores various techniques and methodologies employed in these systems, including Optical Character Recognition (OCR), machine learning, deep learning, and Natural Language Processing (NLP) approaches. The review discusses the challenges and limitations faced by these systems and their potential applications in digitising and preserving handwritten historical documents. It aims to guide future research and development efforts in the domain of Automated Annotation Systems for Document Text Images.

**Keywords:** Automated Annotation . Optical Character Recognition . Handwritten Text Recognition . CNN . RNN . LSTM . CTC . Transformers

#### **Introduction:**

Optical Character Recognition (OCR) is a pivotal technology that enables the transformation of printed or handwritten text from physical documents or images into machine-readable and editable text. It plays a crucial role in converting books, articles,

historical records, and various textual materials into digital formats. There have been numerous remarkable advancements in the field of OCR[22].

While OCR primarily focuses on printed text[21], HTR takes the recognition of text a step further by specialising in handwritten content. [23] introduces an optimised OCR system tailored for handwritten Marathi text document classification and recognition. Many historical documents, archives, and manuscripts contain handwritten text, which presents unique challenges for recognition due to variations in writing styles, ink fading, and paper degradation. HTR technology employs neural networks, machine learning algorithms, and pattern recognition techniques to decipher and transcribe handwritten content, preserving the authenticity and historical significance of such documents. Its primary aim is to convert handwritten text into digital format[18].

Handwritten documents have been a vital source of historical and cultural information, but the task of annotating and transcribing them manually is both time-consuming and error-prone. Automated Annotation prevails over manual annotation due to its superior efficiency and user-friendly nature.

Automated Annotation Systems[20] are used to streamline and improve various tasks related to text analysis, information retrieval, and data organisation. It enhances the organisation, searchability, and understanding of textual data. They typically employ techniques from Natural Language Processing(NLP), machine learning, and data mining to analyse and understand text, enabling the automated

## ii)Result:

### De-GAN and CRAFT Modules:

మిథ నంద్రాల్ ఇంటి చలనం  
బీర్యాగ ఉంటుంది. ఇంటి కదలకు  
నోధించే ఇల్ల గురువులుం మీద  
ఎక్స్‌వగా, సుసుఫ్ లెలుం మీద  
తెక్కువగా టస్టు కనిపిస్తుంది.  
ఒంటి ఇసుక మీద కంచే  
సుసుని పొలరాట గుమ్మ మీద

Original Image

మిథ నంద్రాల్ ఇంటి చలనం  
బీర్యాగ ఉంటుంది. ఇంటి కదలకు  
నోధించే ఇల్ల గురువులుం మీద  
ఎక్స్‌వగా, సుసుఫ్ లెలుం మీద  
తెక్కువగా టస్టు కనిపిస్తుంది.  
ఒంటి ఇసుక మీద కంచే  
సుసుని పొలరాట గుమ్మ మీద

Image After De-GAN

మిథ నంద్రాల్ ఇంటి చలనం  
బీర్యాగ ఉంటుంది. ఇంటి కదలకు  
నోధించే ఇల్ల గురువులుం మీద  
ఎక్స్‌వగా, సుసుఫ్ లెలుం మీద  
తెక్కువగా టస్టు కనిపిస్తుంది.  
ఒంటి ఇసుక మీద కంచే  
సుసుని పొలరాట గుమ్మ మీద

Image After CRAFT

Fig 9: Results generated