
MANTHAN
राष्ट्रीय सुरक्षा सर्वोपरि
Solution By: Prose Radars
PSID: INTL-IVA-14
Team Leader Name: Rupin

Idea Introduction

Hindi Text OCR

Tech Stack	Resources Used	Any Third-Party API/Services used
1.CNN	1.Tensorflow	1. Google Vision
2.NLP	2. NLTK	2. Firebase
3.OpenCV	3.flask	
	4.firebase	

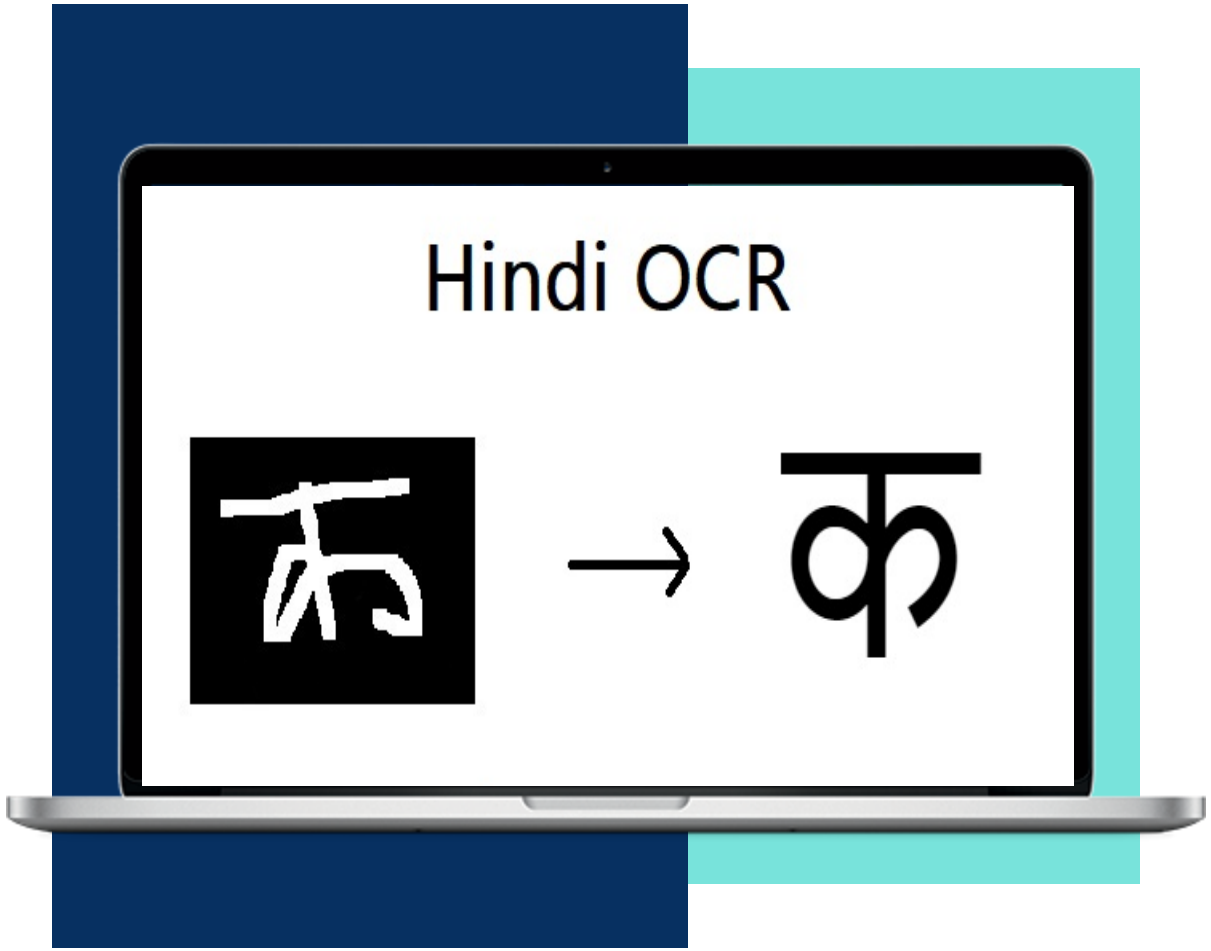
Handwritten Hindi characters to digitized(UTF-8) characters

We shall be implementing a Neural network-based character recognition model which shall be able to match each hand-written characters to its UTF-8 equivalent(in Hindi or English) and hence, digitize the entire document without changing the format.

Your Approach Towards Idea

MANTHAN
राष्ट्रीय सुरक्षा सर्वोपरि

Azadi Ka Amrit Mahotsav



Spiral Model of Software Development

Spiral Model is chosen for to develop the software as it allows us to make small changes and test them out to build a robust software with minimal errors/faults

Most of the existing OCR software do not provide character recognition and those which do , it's available for English only

We plan to implement character by character recognition to allow them to be easily edited and retrieved

The software should be able to generate the digitized copy in the same format as the original one

Team Slide

MANTHAN
राष्ट्रीय सुरक्षा सर्वोपरि

Azadi Ka Amrit Mahotsav

Member 1
Rupin Patel

Mentor
Dr. Varalakshmi M

Member 2
Shreya Kundu

Member 3
Parth Shah

Member 4
Aatman Prajapati

Member 5
Nishit Pandit

Development Pipeline

MANTHAN
राष्ट्रीय सुरक्षा सर्वोपरि

Azadi Ka Amrit Mahotsav

Step A

Proposed Development Plan

1.Find the suitable Dataset
2.Design the feasible neural network
3.Testing
4.Performance optimization.

Step B

Documents to be prepared

1.Proposal Document
2. Performance description
3. Usage Instructions

Step C

Repository Link

- Github Link

Step D

Third-Party APIs used

1. Google Vision
2. Firebase

Step E

Proposed UI and functional flow

1. Upload document to website
2. Click to process the document
3.View the generated document
4. Download the document as PDF

MANTHAN
राष्ट्रीय सुरक्षा सर्वोपरि

Azadi Ka Amrit Mahotsav

Vision of Innovation/Idea/Solution

We started by looking into the existing software which exist for OCR recognition in Hindi, most of them just scanned the copy and just gave an image of the document with highlighted text

In our solution we propose a character-based recognition which shall convert the document to UTF-8 encoded characters which shall help in increasing the readability as well usability of the document in hand.

We plan to design a workable model by the end of the competition with at-least 85-90% accuracy. We shall then work on to increase the accuracy up to 95% and then release a stable model.

We found out that using a character recognition can be performed more accurately and faster via VCG16 library instead of using EasyOCR library.