Mini Project
Final Presentation

Instructor : Dr. Pavan Chakraborty

Handwritten and Machine Printed Text Detection

Table of contents

Our team

Mrityunjaya Tiwari
IIT2019239

Raunak Rathore
IIT2019222

Amanjeet Kumar
IIB2019239

Jyoti Verma
IIT2019202

01

About the project

Introduction

- The presence of printed and handwritten text in the same image of the document poses considerable problems as each mode requires different processing to recognize the corresponding characters.
- Users need to click pictures of their pic/doc that has both printed and Handwritten text, and upload that to our webpage.
- The system will analyse uploaded picture and give output with classification printed texts (show in blue color), handwritten texts (show in green color) and non-texts (show in yellow color).

02

The Proposed Approach

Approach

The proposed methodology consists of two stages.

- The first stage is to localize the possible text regions from the document images, and
- The second stage is to classify the localized portions of the image as handwritten, printed, non-text or in a few cases, mixed/combined text using the features extracted from the images of word or word-like segments.

Approach

Step1

The image is first converted into a single channel grayscale image

Step2

Now Otsu's binarization is performed on the output image

Step3

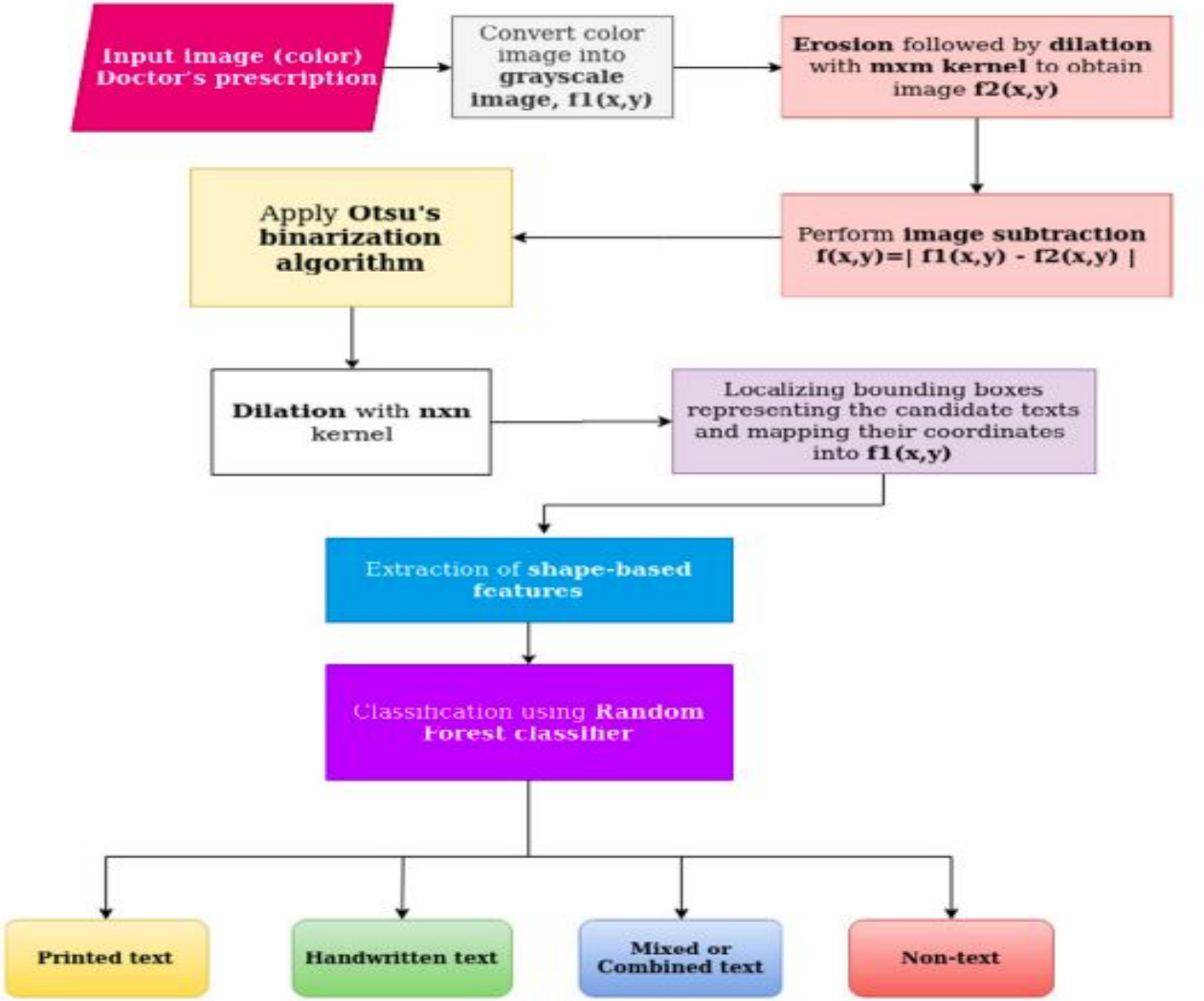
Now, the bounded box of each of the patches is determined and these parts are cropped out from the original image

Step4

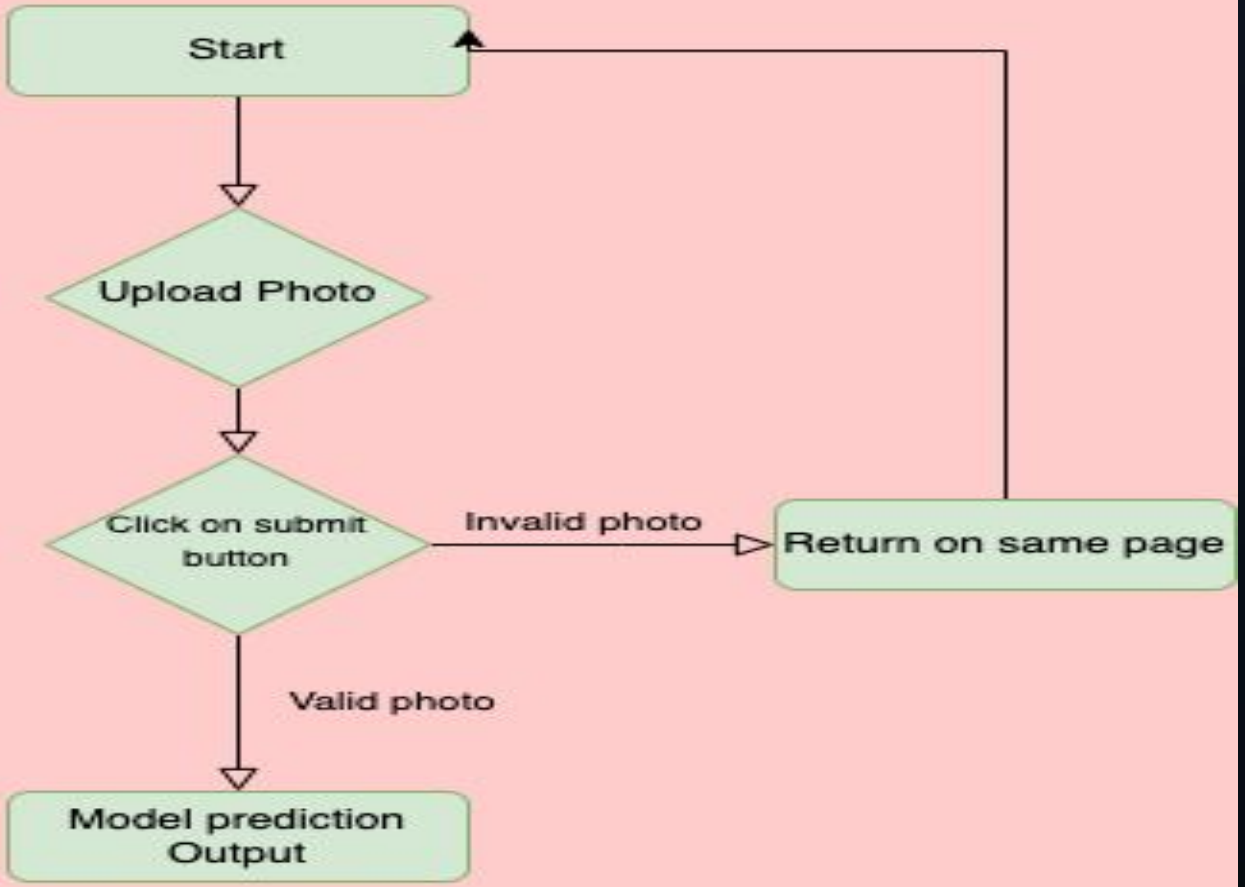
Training the model. Deployment on flask server.

03

Flow Chart

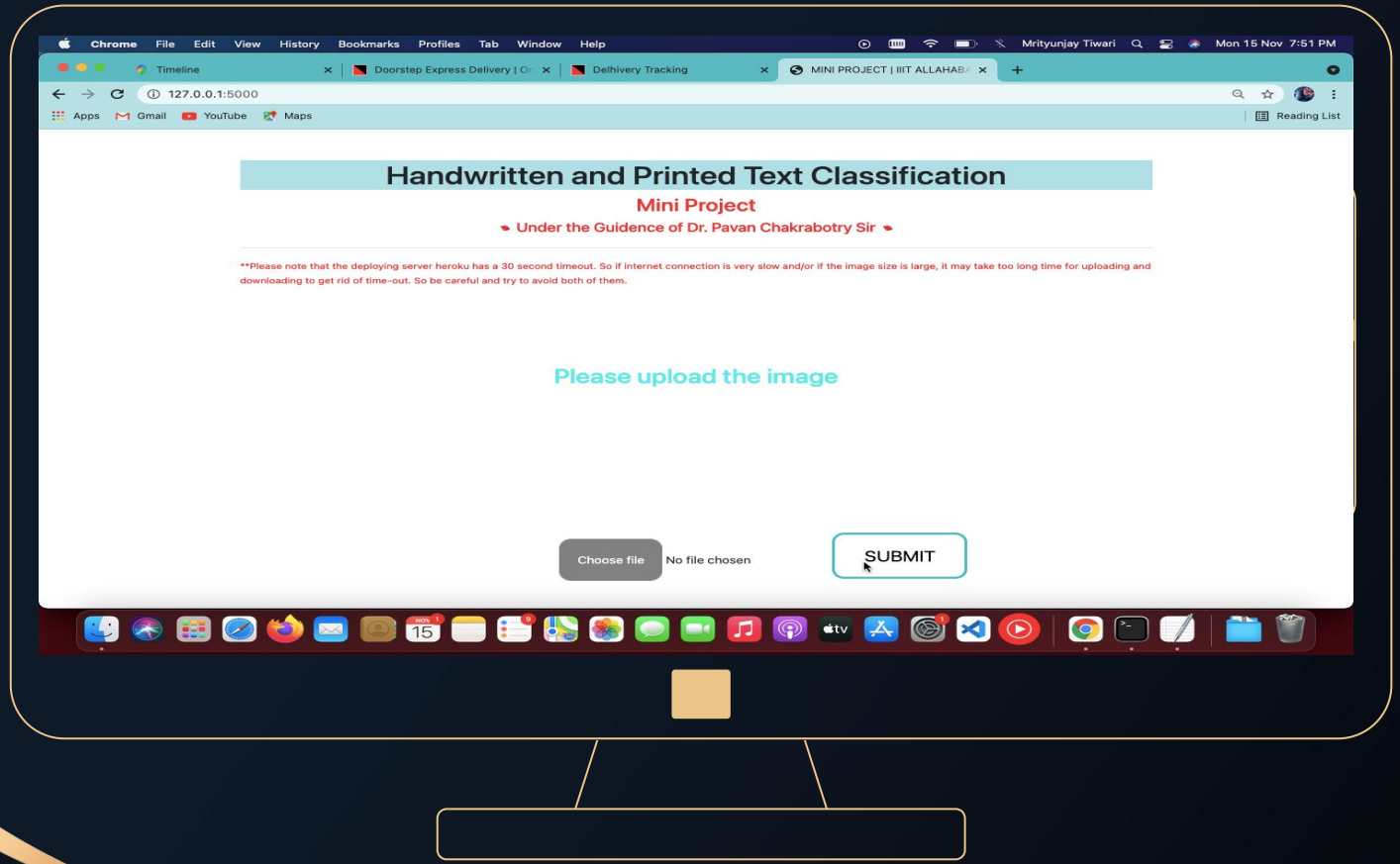

Input image (color) Doctor's prescription
Convert color image into grayscale image, f1(x,y)
Erosion followed by dilation with mxm kernel to obtain image f2(x,y)
Perform image subtraction f(x,y)=| f1(x,y) - f2(x,y) |
Apply Otsu's binarization algorithm
Dilation with nxn kernel
Localizing bounding boxes representing the candidate texts and mapping their coordinates into f1(x,y)
Extraction of shape-based features
Classification using Random Forest classifier
Printed text
Handwritten text
Mixed or Combined text
Non-text

Flow Chart of Deployed Flask Model

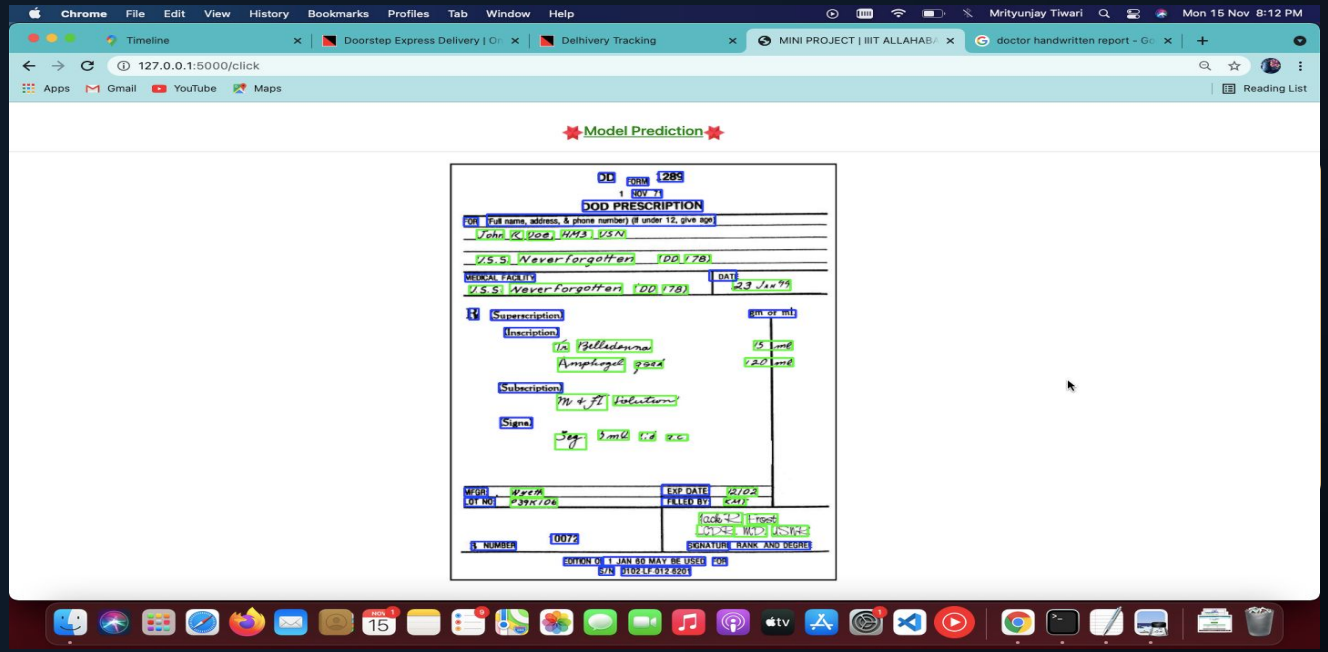


04
Results and discussion

Website


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Mrityunjay Tiwari
Mon 15 Nov 7:51 PM
Timeline
Doorstep Express Delivery | Or
Delhivery Tracking
MINI PROJECT | IIIT ALLAHAB
127.0.0.1:5000
Apps
Gmail
YouTube
Maps
Reading List
Handwritten and Printed Text Classification
Mini Project
Under the Guidence of Dr. Pavan Chakrabotry Sir
**Please note that the deploying server heroku has a 30 second timeout. So if internet connection is very slow and/or if the image size is large, it may take too long time for uploading and downloading to get rid of time-out. So be careful and try to avoid both of them.
Please upload the image
Choose file
No file chosen
SUBMIT

Model Prediction



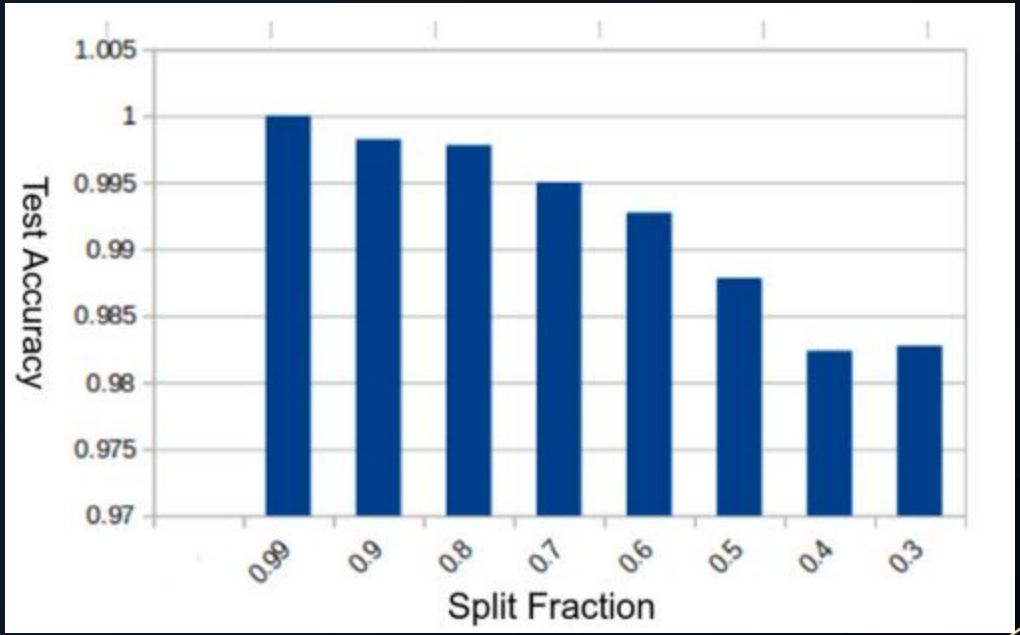
Predicted results

97.72%

Model Accuracy

2.28%

Loss



05

Conclusion

Conclusion

- In this project, a method has been proposed to Classify handwritten and machine printed text present in the same image & according to the text, printed texts (will show in blue color), handwritten texts (will show in green color) and non-texts (will show in yellow color).
- As the proposed method has successfully classified the printed and handwritten texts in the documents and with a very low complexity, this can easily be embedded with a recognition module as an additional resource requirement.

06

References

Resources

- V.Pal and B.B.Chaudhuri, “Machine-printed and handwritten text lines identification”, Pattern Recognition Letters, 22, 2001, pp.431-441.
- https://searchcontentmanagement.techtarget.com/definition/OCR-optical-character-recognition
- How to create salt and pepper noise in an image. https://www.projectrhea.org/rhea/index.php/How to Create Salt and Pepper Noise in an Image
- Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Thanks!