MINI PROJECT

C2 PRESENTATION

Instructor : Dr. Pavan Chakraborty

Handwritten and Machine Printed Text Detection

Group Members:

Mrityunjay Tiwari - IIT2019239
Raunak Rathore - IIT2019222
Amanjeet kumar - IIB2019006
Jyoti Verma - IIT2019202

Content

- INTRODUCTION
- THE PROPOSED APPROACH
- FLOW CHART
- SCREENSHOTS OF LIVE SYSTEM
- TIMELINE
- CONCLUSION
- RESULTS
- REFERENCES

Introduction

- Nowadays, one can observe a rapidly growing number of digitization initiatives in libraries and archives, involving a variety of document types.
- The presence of printed and handwritten text in the same document image gives rise to significant issues since each modality requires different treatment to recognize the corresponding characters.
- So, it is necessary to separate the machine printed and handwritten text before applying different recognition methodologies to each.

THE PROPOSED APPROACH

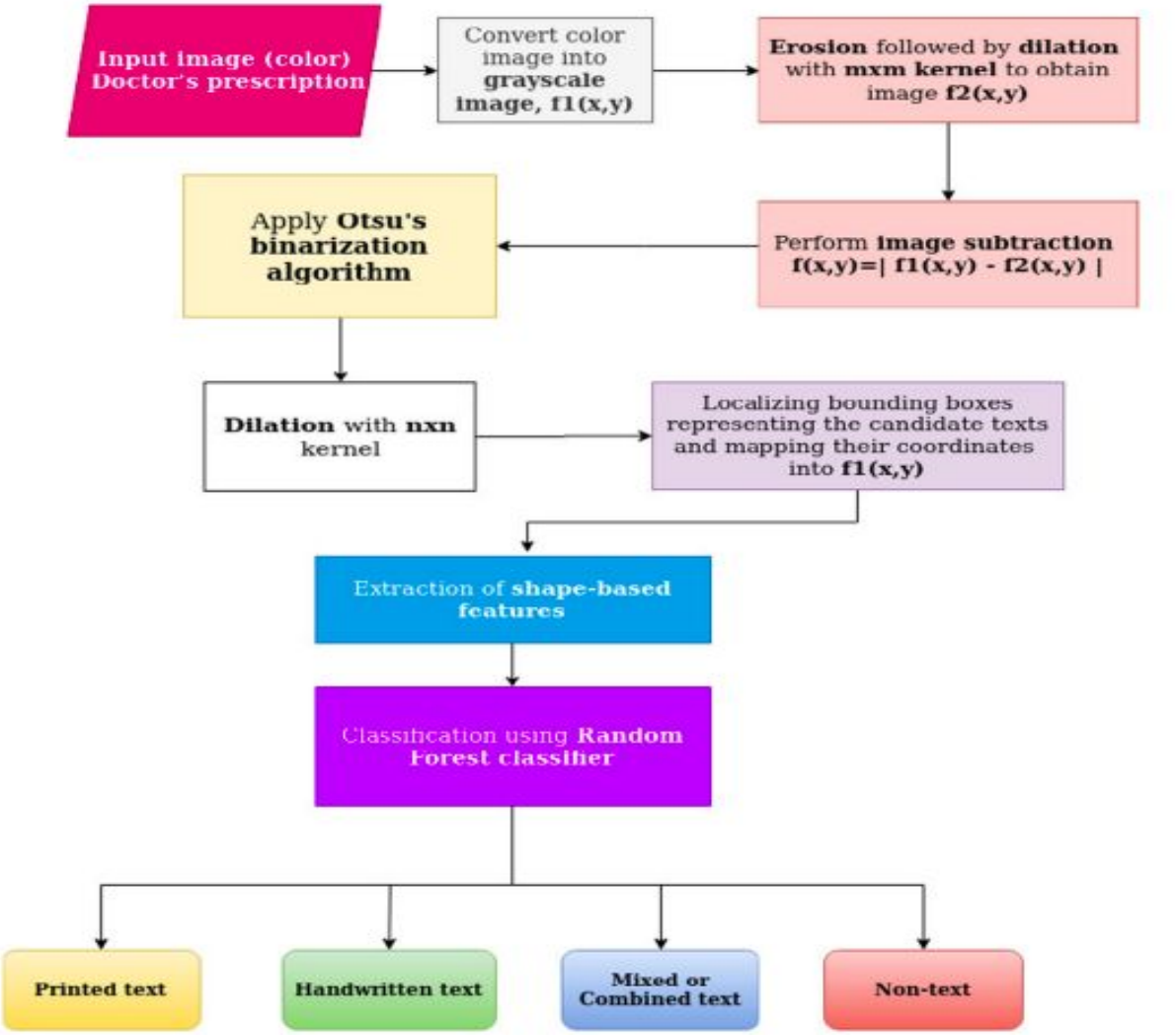
- **OCR (optical character recognition)** is the use of technology to distinguish printed or handwritten text characters inside digital images of physical documents, such as a scanned paper document.
- The basic process of OCR involves examining the text of a document and translating the characters into code that can be used for data processing.

Optical Character Recognition

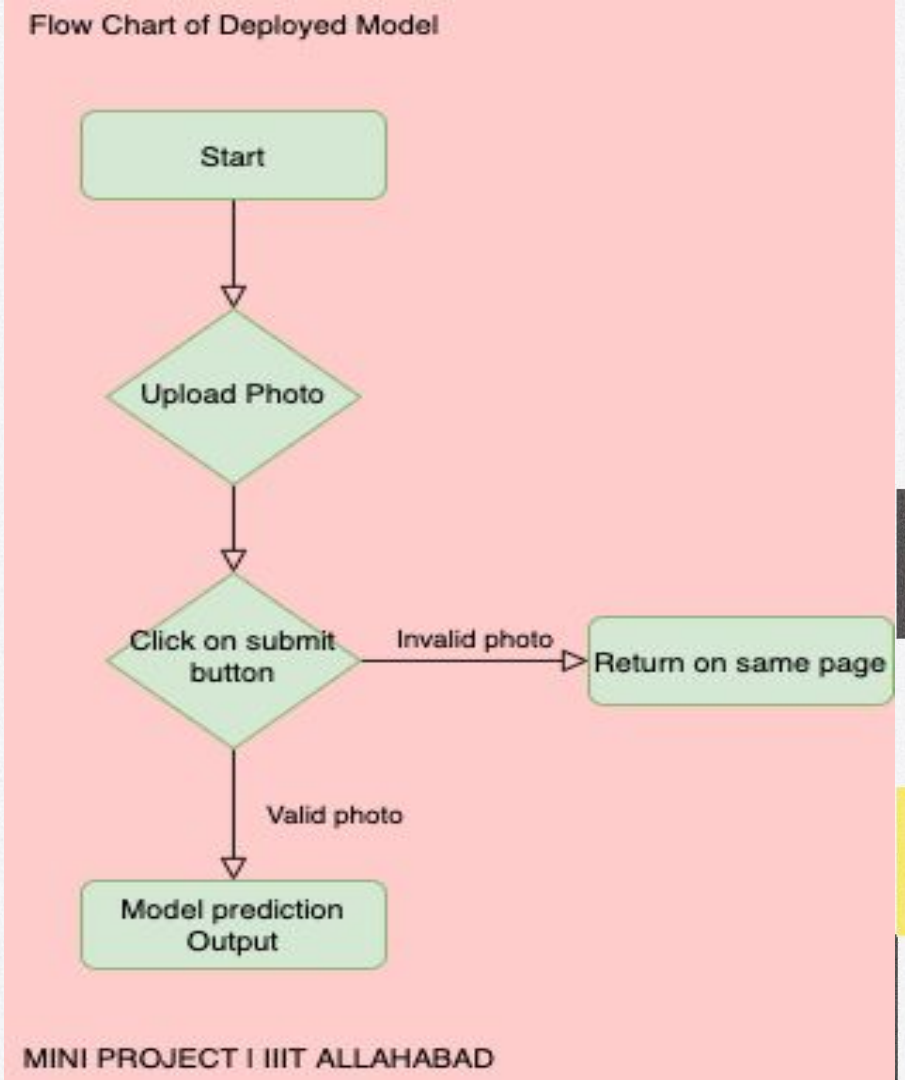
The proposed methodology consists of two stages.

- Localize the possible text regions from the documents
- Then classify the localized portions of the image as handwritten, printed, non-text or in a few cases, mixed/combined text

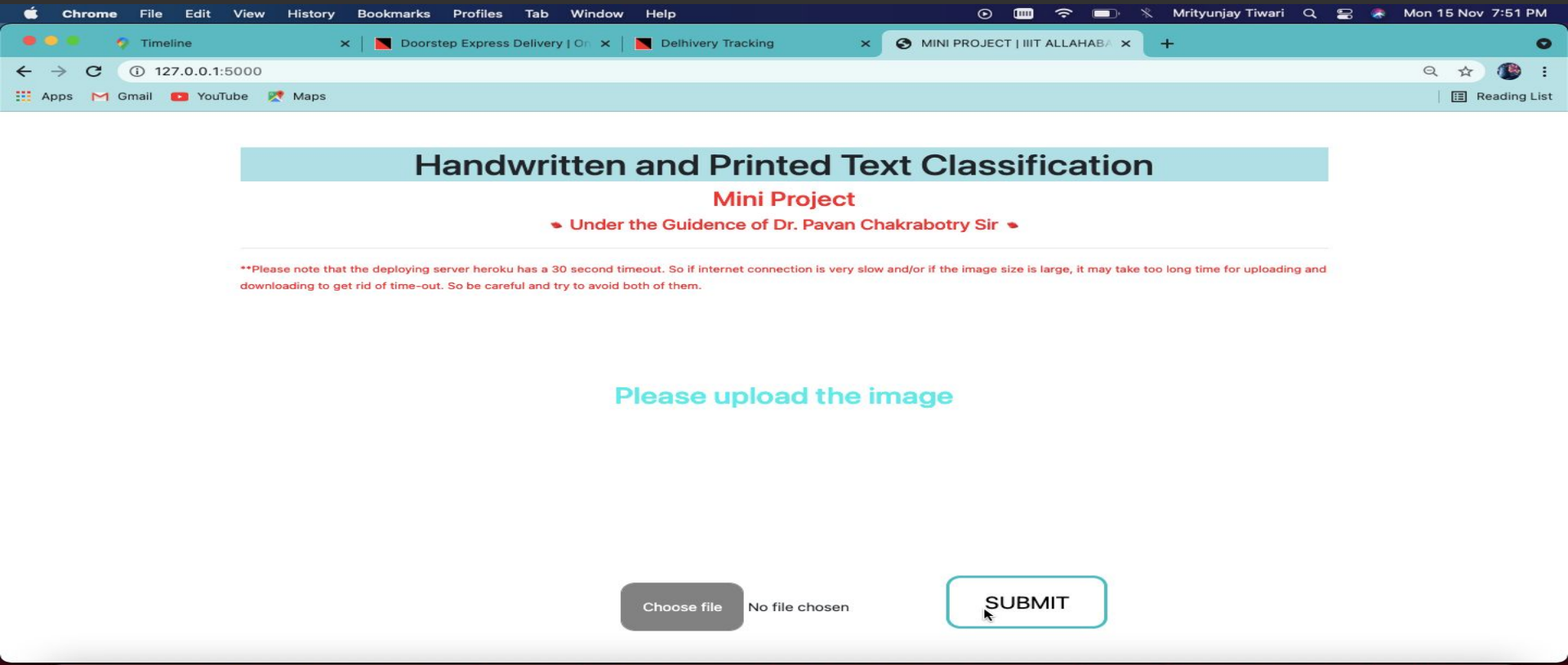
FLOW CHART OF PROPOSED METHOD



FLOW CHART OF DEPLOYED MODEL



Screenshots of deployed model



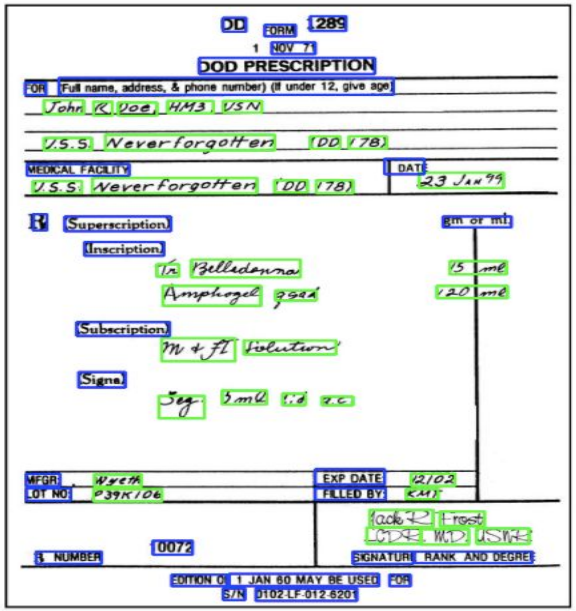
Chrome File Edit View History Bookmarks Profiles Tab Window Help — Mrityunjay Tiwari — Mon 15 Nov 8:12 PM

Timeline | Doorstep Express Delivery | On | Delhivery Tracking | MINI PROJECT | IIIT ALLAHABA | doctor handwritten report - Go

127.0.0.1:5000/click

Apps Gmail YouTube Maps — Reading List

Model Prediction



Output of our trained model

TIMELINE

1.Work done so far:

- Selected a specific approach to solve the given problem of Handwritten and Machine Printed Text Detection.
- Selected a dataset for the given problem.
- Till now, we have deployed our model on our local system.It runs properly with good accuracy and in very less efficient time . Now, we are working to deploy our model in live servers like AWS, Heroku .

2.Future work :

- Host on live server (like Heroku, AWS, azuse, Google cloud etc...)

Conclusion

- In this project, a method has been proposed to classify printed and handwritten texts in documents.
- 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.
- The scope of the present work can be used for direct application by the researchers and the netizens according to their requirements.

References

- 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

THANK YOU