OPTICAL CHARACTER RECOGNITION

BANGLA LANGUAGE

Team Cybernauts.

Bangladesh University of Business and Technology [BUBT]

Members:

Johirul Islam Tutul (19202103001)

Mahanur Alam (19202103030)

Nabila Rahman (19202103056)

Tanmoy Islam Opu(19202103026)

Tanvir Ahmed (19202103424)

Supervised by:

Khan Md. Hasib

Assistant Professor

BUBT

Committed to Academic Excellence

BANGLADESH UNIVERSITY OF BUSINESS AND TECHNOLOGY

CONTENTS

- Introduction
- Applications
- CNN based Model
- Deep learning for OCR
- Tools
- Motivation
- Challenges of OCR for Bangla Language
- Conclusion

INTRODUCTION

Optical Character Recognition (OCR) is a technology that enables the recognition of text in digital images or scanned documents. It is a process of converting printed or handwritten text into machine-encoded text. OCR has been used to convert many languages into machine-readable forms, including Bangla.

23-07-2023

Application of OCR Bangla Language

1. Digitalizing Bangla books and documents
2. Extracting text from Bangla images
3. Creating searchable Bangla archives
4. Automating data entry in Bangla language
5. Generating Bangla subtitles for videos and movies

23-07-2023

CNN BASED MODEL

1. Convolutional Neural Network (CNN)
2. Preprocessing
3. Image recognition
4. Noise reduction
5. Extraction feature
6. Character recognition

23-07-2023
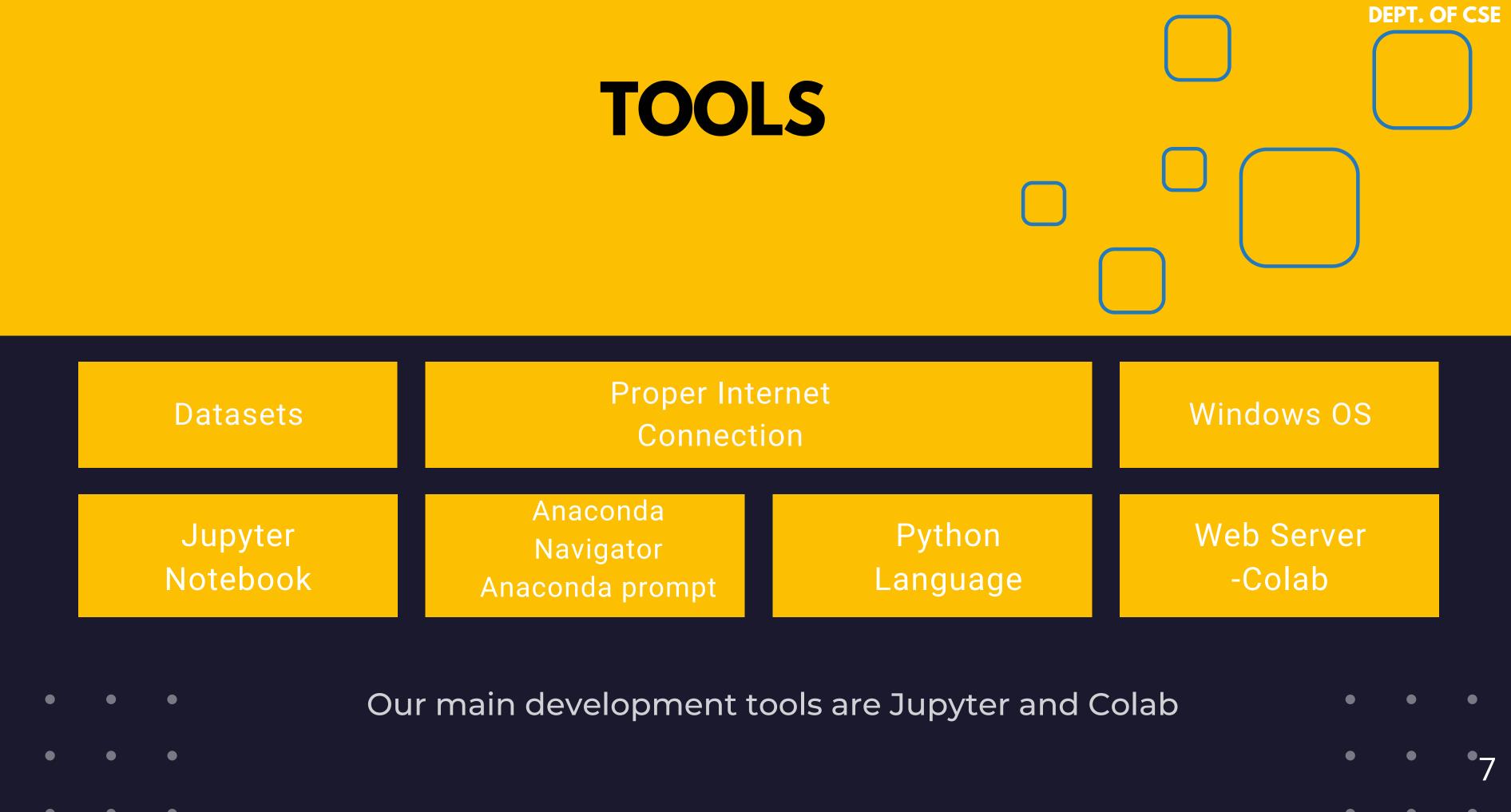
DEEP LEARNING IMPLEMENTATION

- **ARTIFICIAL INTELLIGENCE (AI)**
- **NEURAL NETWORKS**
- **DEEP NEURAL NETWORKS (DNN)**
- **CONVOLUTIONAL NEURAL NETWORKS (CNN)**
- **RECURRENT NEURAL NETWORKS (RNN)**
- **TRAINING DATA**

23-07-2023

TOOLS

Datasets	Proper Internet Connection	Windows OS
Jupyter Notebook	Anaconda Navigator Anaconda prompt	Python Language	Web Server -Colab

Our main development tools are Jupyter and Colab

23-07-2023

Benefits for OCR Bangla Language

🗎 **Language Specialist**

⚙ **Book publisher**
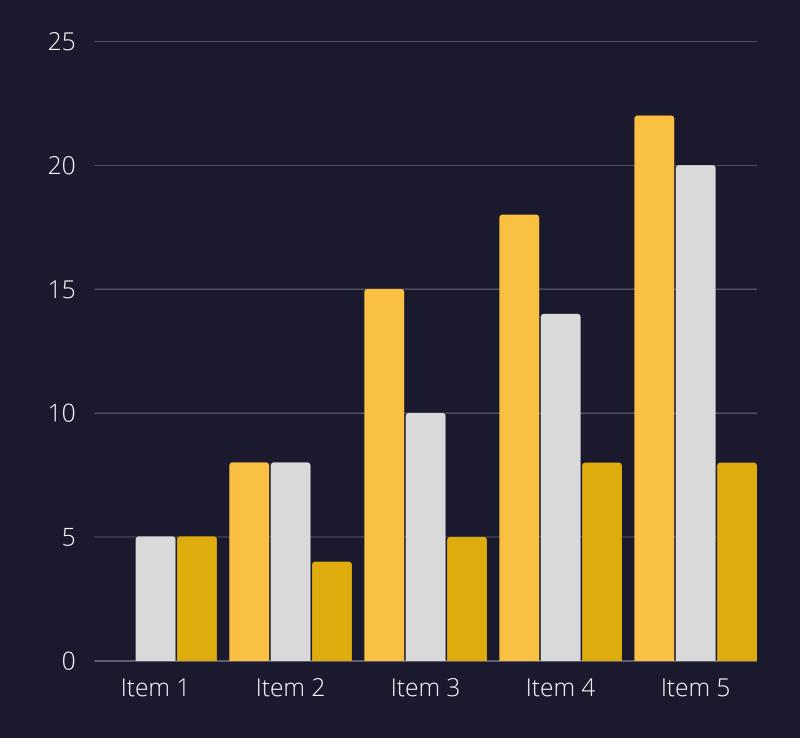
⭐ **Author**

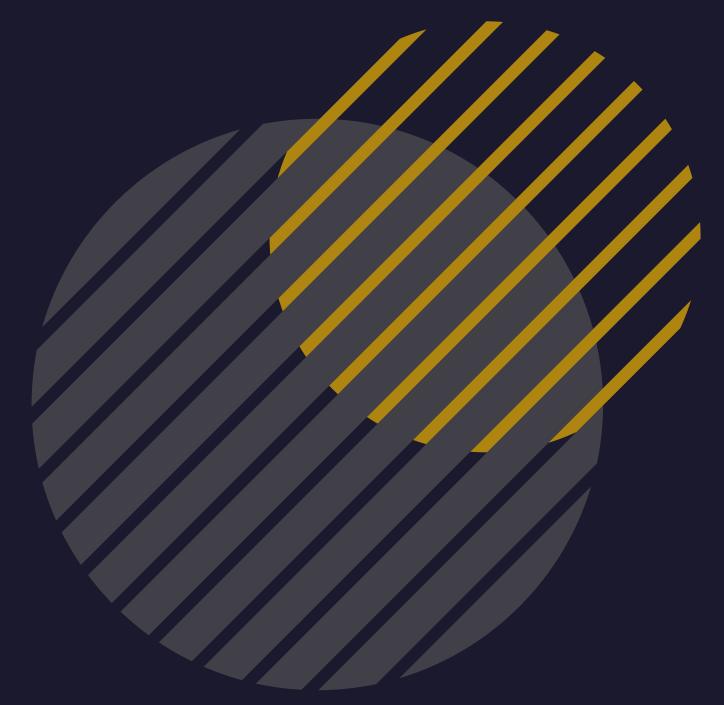💡 **Develope Transport management.**

⭐ **Develope Transport management.**

8

23-07-2023

Challenges

1. Lack of training data: The accuracy of an OCR engine depends heavily on the quality and quantity of training data, and there may be a lack of high-quality training data available for Bangla OCR.
2. Limited resources: Many OCR engines require significant computational resources, and these resources may not be readily available in some Bangla-speaking regions.
3. Vocabulary and language complexity: Bangla language has a complex vocabulary and grammatical rules, which can make it difficult for OCR engines to accurately recognize and understand text.
4. Character segmentation: Separating individual characters in Bangla text can be challenging due to the presence of diacritic marks and ligatures.

23-07-2023

THANKS
FOR WATCHING

ANY QUESTION?