Aadhaar Card Information Extraction Project Overview

In this project, I developed a robust solution to extract essential information from Aadhaar card images, including the Aadhaar number, name, date of birth, and gender, using Optical Character Recognition (OCR) and image processing techniques. The main goal was to automate the data extraction process, ensuring accuracy while handling different types of Aadhaar cards, including PVC Aadhaar and eAadhaar.

# Key Features:

1. Image Preprocessing:

- Used OpenCV to preprocess images, including converting them to grayscale and applying thresholding techniques for better OCR performance.  
 - Implemented cropping techniques to handle the unique layout of eAadhaar cards, specifically focusing on regions containing relevant information.

2. Optical Character Recognition (OCR):

- Utilized Tesseract OCR to extract text from the preprocessed images. Custom configurations and pattern matching (regular expressions) were applied to isolate key details such as the Aadhaar number, name, and date of birth.  
 - Improved accuracy by cleaning and formatting the extracted text using regular expressions and structured search patterns.

3. Aadhaar Validation and Format Identification:

- Integrated two convolutional neural network (CNN) models:  
 1. The first model validates whether the image is a valid Aadhaar card.  
 2. The second model identifies whether the image is a PVC Aadhaar or an eAadhaar.  
 - Based on the format identified, appropriate preprocessing steps, such as image cropping, are applied before text extraction.

4. Tools & Technologies:

- Libraries: OpenCV, Tesseract OCR, TensorFlow/Keras for deep learning models.  
 - Techniques: Image preprocessing, CNN for classification, OCR for text extraction.  
 - Model Deployment: Developed and tested the solution on local systems using pre-trained models.

# Outcomes:

The project successfully automated the extraction of essential Aadhaar details from both PVC and eAadhaar images. The system is designed to handle variations in image quality and format, ensuring minimal manual intervention and achieving high accuracy in data extraction.