**Assignment No: 5**

**Problem Statement:-**

Text identification using **OpenCV**, **Tesseract (OCR)**, and a deep neural network.

**Theory:-**

Optical Character Recognition (OCR) is the process of identifying and extracting text from images. **Tesseract** is a popular OCR engine, and **OpenCV** is used for pre-processing images to improve OCR performance. A deep neural network can be used to enhance accuracy or identify specific patterns in the text.

**Methodology:-**

1. **Image Preprocessing**:
   * Use **OpenCV** to preprocess images, including binarization, noise reduction, and edge detection, to prepare them for OCR.
   * Techniques such as **Gaussian blur** or **thresholding** can improve OCR accuracy.
2. **Text Extraction**:
   * Use **Tesseract OCR** to extract text from the preprocessed images.
   * Train a deep neural network if additional pattern recognition or filtering is needed for the text.
3. **Post-Processing**:
   * Apply language models or rules to correct errors and refine the extracted text.
4. **Applications**:
   * This can be applied to scenarios like automatic number plate recognition (ANPR), document digitization, and form processing.

**Conclusion:-**

We implemented text identification using OpenCV for image preprocessing and Tesseract for OCR, achieving accurate text extraction from images.