

**INTERNSHIP ON**  
**Detection of Round Seals and Handwritten Dates on Medical Certificates**  
**UNDER**  
**CENTER FOR DEVELOPMENT OF IMAGING TECHNOLOGY(C-DIT)**  
**- TRIVANDRUM**

**TEAM MEMBERS**  
**Arjun S Nair (TKM22CS039)**  
**Aromal RM (TKM22CS042)**  
**Krishnanand PS (TKM22CS093)**

**BACHELOR OF TECHNOLOGY**  
**In**  
**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**  
**THANGAL KUNJU MUSALIAR COLLEGE OF ENGINEERING**  
**KARIKODE - 691005**

## **Problem Statement**

Automate the detection of round seals and extraction of handwritten dates from scanned or photographed medical certificates to streamline verification processes.

## **Introduction**

This project focuses on leveraging computer vision (CV) and Optical Character Recognition (OCR) techniques to detect round seals and recognize handwritten dates in medical certificates. The solution aims to improve efficiency and accuracy in verifying certificates, contributing to better automation in administrative processes.

## **System Workflow Details**

- **Data Collection:** Collect a diverse dataset of medical certificates with variations in image quality, layout, seal design, and handwritten text styles. Use tools like LabelImg to mark regions containing seals and handwritten dates.
- **Image Preprocessing:** Convert images to grayscale, reduce noise using Gaussian blur, and enhance edges using Canny edge detection. Extract potential regions containing seals and handwritten text.
- **Round Seal Detection:** Use Hough Circle Transform for initial detection and train a YOLO-based model for precise seal identification.
- **Handwritten Date Recognition:** Apply OCR (e.g., Tesseract) to extract text from the region of interest. Validate dates using regex to ensure formats like dd/mm/yyyy.
- **Pipeline Integration:** Combine preprocessing, detection, and recognition into an automated pipeline.
- **Testing and Validation:** Evaluate seal detection accuracy (precision, recall) and measure date recognition accuracy using ground truth comparisons.

## **System Workflow Timeline**

### **Problem Identification and Study**

**Duration:** 3 weeks

**Start Date:** October 26, 2024

**Key Accomplishments:**

- Defined the problem of detecting seals and extracting dates.
- Studied challenges related to image quality, handwritten text, and shape detection.
- Reviewed traditional CV and modern deep learning methods.

### **Selection of Techniques and Tools**

**Duration:** 2 weeks

**Start Date:** November 16, 2024

**Key Accomplishments:**

- Selected Hough Circle Transform for initial seal detection.
- Chose Tesseract OCR for handwritten date recognition.
- Designed preprocessing pipeline with grayscale conversion, noise reduction, and edge enhancement.

## **Data Collection and Preparation**

**Duration:** 3 weeks

**Start Date:** November 30, 2024

**Key Accomplishments:**

- Annotating a sample dataset of certificates for testing workflows.
- Preparing pipelines for cropping and processing regions of interest (seals and dates).

## **Model Training and Integration**

**Duration:** 4 weeks

**Start Date:** December 21, 2024

**Key Accomplishments:**

- Gather the full dataset with labeled seals and date annotations.
- Organize data for training and testing.
- Train the YOLO model for seal detection and validate Tesseract OCR for dates.
- Optimize hyperparameters to improve accuracy.
- Evaluate seal detection and date extraction accuracy against the ground truth.

## **Testing and Finalization**

**Duration:** 2 weeks

**Start Date:** January 18, 2025

**Key Accomplishments:**

- Integrate the pipeline and evaluate its performance.
- Prepare documentation and finalize the solution for deployment.