

OCR FOR KYC DOCUMENTS

A PROJECT REPORT

Submitted by

INFANT VINOTH P (20C033)

Interning with the Company

**CHOLAMANDALAM INVESTMENT AND FINANCE
COMPANY LTD.**

in partial fulfilment for the award of the degree

Of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING



THIAGARAJAR COLLEGE OF ENGINEERING, MADURAI – 15

(A Government Aided Autonomous Institution Affiliated to Anna University)

ANNA UNIVERSITY: CHENNAI 600 025

MAY 2024

THIAGARAJAR COLLEGE OF ENGINEERING MADURAI-15

(A Government Aided Autonomous Institution Affiliated to Anna University)

BONAFIDE CERTIFICATE

Certified that this project report “**OCR for KYC Documents**” is the bonafide work of “**INFANT VINOTH (20C033)**” who carried out the project work under my supervision during the Academic Year 2023-2024.

SIGNATURE

Dr.S.MERCY SHALINIE ,M.E., Ph.D.,

HEAD OF THE DEPARTMENT

COMPUTER SCIENCE AND
ENGINEERING

THIAGARAJAR COLLEGE OF ENGG,
MADURAI-625 015

SIGNATURE

Dr.D.NAGENDRA KUMAR

ASSISTANT PROFESSOR

COMPUTER SCIENCE AND
ENGINEERING

THIAGARAJAR COLLEGE OF ENGG,
MADURAI-625 015

Submitted for the VIVA VOCE Examination held at Thiagarajar College of

Engineering on

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

I would like to thank my supervisor **Mr.Arun.R.S** of **Cholamandalam Investment and Finance Company Limited** under whom I am pursuing my industrial internship for his continued support and guidance. The internship was a fruitful experience where I was able to learn a lot of new things and get a sight into how the industry really is.

I wish to express my deep sense of gratitude to **Dr. M. Palaninatharaja**, Principal of Thiagarajar College of Engineering for his support and encouragement throughout this project work.

I wish to express my sincere thanks to **Dr. S Mercy Shalinie** , Head of the Department of Computer Science and Engineering for her support and ardent guidance.

I owe my special thanks and gratitude to **Dr.D.Nagendra Kumar** , Head of the Department of Computer Science and Engineering for her guidance and support throughout our project.

I am also indebted to all the teaching and non teaching staff members of our college for helping us directly or indirectly by all means throughout the course of our study and project work.

I extremely thank my parents, family members and friends for their moral support and encouragement for my project.

ABSTRACT

The OCR project for Aadhaar cards is designed with the overarching goal of reducing reliance on external APIs such as Karza, thus enhancing data privacy and reducing operational costs. By integrating the YOLOv8 model at the frontend, the system gains advanced object detection capabilities, enabling it to accurately locate and extract Aadhaar card images from webcam input with minimal user intervention. The YOLOv8 model's real-time processing capabilities ensure efficient handling of image streams, further enhancing user experience and system responsiveness. Moreover, the project incorporates robust security measures to safeguard sensitive Aadhaar card information, including encryption protocols and secure data transmission methods. Through these features, the system not only streamlines the OCR process but also prioritises data privacy and security, aligning with regulatory requirements and user expectations. Additionally, the project emphasises scalability and extensibility, allowing for future integration with other identity documents and expansion into diverse use cases beyond KYC verification. In addition to its core functionality of Aadhaar card OCR, the project also masks the Aadhaar number in the provided Aadhaar image as the Aadhaar image must not be stored with the Aadhaar number visible in the company's server as per the RBI rules. The integration of machine learning algorithms not only enhances accuracy in text recognition but also enables adaptive learning capabilities, allowing the system to continuously improve its performance over time through feedback mechanisms and dataset enrichment. Furthermore, the project incorporates a modular architecture, facilitating easy integration with existing software systems and enabling customization to meet specific business requirements.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	4
	LIST OF TABLE	5
	LIST OF FIGURES	6
	LIST OF ABBREVIATIONS	7
1	INTRODUCTION	1
1.1	INTRODUCTION ABOUT THE COMPANY PLACED	1
1.2	INTRODUCTION OF THE PROJECT	2
1.3	PROBLEM STATEMENT	3
1.4	OBJECTIVE	4
2	LITERATURE REVIEW	6
3	SOFTWARE REQUIREMENTS	9
3.1	SOFTWARE SPECIFICATIONS	9
4	PROPOSED SYSTEM	18
4.1	METHODOLOGY	18
5	IMPLEMENTATION	25
6	RESULT AND DISCUSSION	31
7	CONCLUSION AND FUTURE WORKS	40

TABLE OF FIGURES

S.No	FIGURE	PAGE NO
1.1	LOGO	1
1.2.	OBJECTIVES	5
3.1	FRONTEND TECHSTACK	10
3.2	BACKEND TECHSTACK	10
3.3	HTML,CSS,JAVASCRIPT	11
3.4	JAVA	12
3.5	SPRING BOOT	12
3.6	EASYOCR	13
3.7	TESS4J	14
3.8	OPENCV	15
3.9	YOLOV8	16
3.10	TENSORFLOW JS	17
4.1	DEVELOPMENT PHASE	24
5.1	FLOWCHART	30
6.1	CARD DETECTION PAGE	32
6.2	GRAYSCALE IMAGE	33
6.3	REDACTED OUTPUT	34
6.4	API RESPONSE	35

LIST OF ABBREVIATIONS

	ABBREVIATION
CIFCL	Cholamandalam Investment And Finance Company Limited
NBFCs	Non-Banking Financial Companies
SMEs	Small and Medium Businesses
CSEL	Consumer & Small Enterprise Loans
SBPL	Secured Business Personal Loans
OCR	Optical Character Recognition
Tess4J	Tesseract For Java
Tensorflow JS	Tensorflow JavaScript
OpenCV	Open Computer Vision
API	Application Programming Interface
SSPL	Server Side Public Licence
REST	Representational State Transfer