

## ABSTRACT

OCR based Aadhar & driving license info Extraction System may be a time period embedded system that mechanically acknowledges the kind of document whether or not it's a Aadhar or driving license and extract the out there info from it. There square measure several applications starting from advanced security systems to common official work. OCR primarily based Aadhar & driving license info Extraction System has advanced characteristics because of various effects like totally different pattern in numerous Aadhar Card, totally different Spacing in text etc. Most of the OCR primarily based info Extraction System square measure designed mistreatment proprietary tools like MATLAB that takes a protracted method and time and conjointly will have many limitations and conjointly they're unable to sight pattern and can't extract the desired info severally. this concept presents an efficient technique of implementing OCR primarily based Aadhar & driving license info Extraction System mistreatment Free software package together with Python and therefore the Open cv Vision Library

## TABLE OF CONTENTS

Chapter. No	Title	Page No.
	<b>Abstract</b>	<b>i</b>
	<b>Table of Contents</b>	<b>ii</b>
	<b>List of Figures</b>	<b>iv</b>
	<b>Abbreviation</b>	<b>v</b>
<b>1.</b>	<b>INTRODUCTION</b>	<b>1-3</b>
	1.1 Introduction	1
	1.2 Motivation	1
	1.3 Problem Definition	2
	1.4 Objective of Project	2
	1.5 Existing System	2
	1.6 Disadvantages of Existing System	2
	1.7 Proposed System	3
	1.8 Advantages of Proposed System	3
<b>2.</b>	<b>REQUIREMENT ANALYSIS</b>	<b>4-5</b>
	2.1 Requirement Analysis	<b>4</b>
	2.2 Requirement Specification	4
	2.2.1 Functional requirement	4
	2.2.2 Non-Functional requirement	4
	2.3 Computational resource requirements	5
	2.3.1 Hardware requirements:-	5
	2.3.2 Software requirements:-	5
<b>3</b>	<b>DESIGN</b>	<b>6-8</b>
	3.1 Introduction	6
	3.2 System Architecture	6
	3.3 UML Diagrams	7
	3.3.1 Class Diagram	7
	3.3.2 Sequence Diagram	8
<b>4</b>	<b>ALGORITHM AND MODULES</b>	<b>9-20</b>

MACHINE LEARNING BASED AADHAR CARD AND DRIVING LICENSE  
DATA EXTRACTION IN DATABASE

	4.1 Algorithm	9
	4.1.1 Update Data Algorithm	9
	4.1.2 Search Data Algorithm	9
	4.2. Flowchart	10
	4.2.1 Update Data Flowchart	10
	4.2.2 Search Data Flowchart	11
	4.3 Modules	11
	4.3.1 Numpy	12
	4.3.2 Pandas	13
	4.3.3 TensorFlow	14
	4.3.4 Opencv-Python	15
	4.3.5 Pytesseract	16
	4.3.6 Nltk	17
	4.3.7 SQLite3	18
<b>5</b>	<b>IMPLEMENTATION</b>	<b>21-31</b>
	5.1 Introduction	21
	5.2 Explanation of Key Functions	21
	5.3 Source Code	21
	5.4 Result Analysis	31
<b>6</b>	<b>TESTING</b>	<b>32-35</b>
	6.1 Overview of Testing	32
	6.2 Types of Testing	32
	6.2.1 Unit testing	32
	6.2.2 Integration testing	32
	6.2.3 Functional testing	33
	6.3 Unit Testing	33
	6.4 Integration Testing	34
	6.5 Acceptance Testing	34
	6.6 Test Cases	34
	6.7 Black Box Testing	35
	6.8 White Box Testing	35
	6.9 Design of Test Cases and Scenarios	35
<b>7</b>	<b>SCREENTHOTS</b>	<b>37-38</b>

MACHINE LEARNING BASED AADHAR CARD AND DRIVING LICENSE  
DATA EXTRACTION IN DATABASE

---

	7.4 Interface Screenshot	37
	7.7.1 Upload Page	37
	7.7.2 Retrieve Page	38
<b>8</b>	<b>CONCLUSION</b>	<b>39</b>
	8.1 Project Conclusion	39
	8.2 Limitation of Project	39
	8.3 Future Enhancements	39
<b>9</b>	<b>REFERENCES</b>	<b>40</b>

## LIST OF FIGURES

Sl No	Title	Page No.
3.1	System Architecture	6
3.2	Class Diagram	7
3.3	Sequence Diagram	8
7.1	Upload Page	37
7.2	Retrieve Page	38

## ABBREVIATION

GUI	Graphical User Interface
PY	Python
OPP	Object Oriented Programming
DRY	Don't Repeat Yourself
PIP	Package Installer for Python
MRO	Method Resolution Order
PEP	Python Enhancement Proposals
BDFL	Benevolent Dictator For Life
REPL	Read-Eval-Print Loop