EXACT SCIENCES

hackerearth

Exact Sciences OCR Hack

Prototype Presentation

Team

TheRenaissance

Introduction

We are the final year students of Computer Engineering from Anand(Gujarat- India). We all have skills in different domains like frontend, backend development and Machine Learning. Here we present our final prototype of Optical Character Recognition Hackathon.

Team **TheRenaissance**

1. **Keyur Khant**
2. **Khushal Gondaliya**
3. **Savan Kansagra**

THE CHALLENGE

EXTRACT FORM FILLED DATA MORE ACCURATE USING OPTICAL CHARACTER RECOGNITION

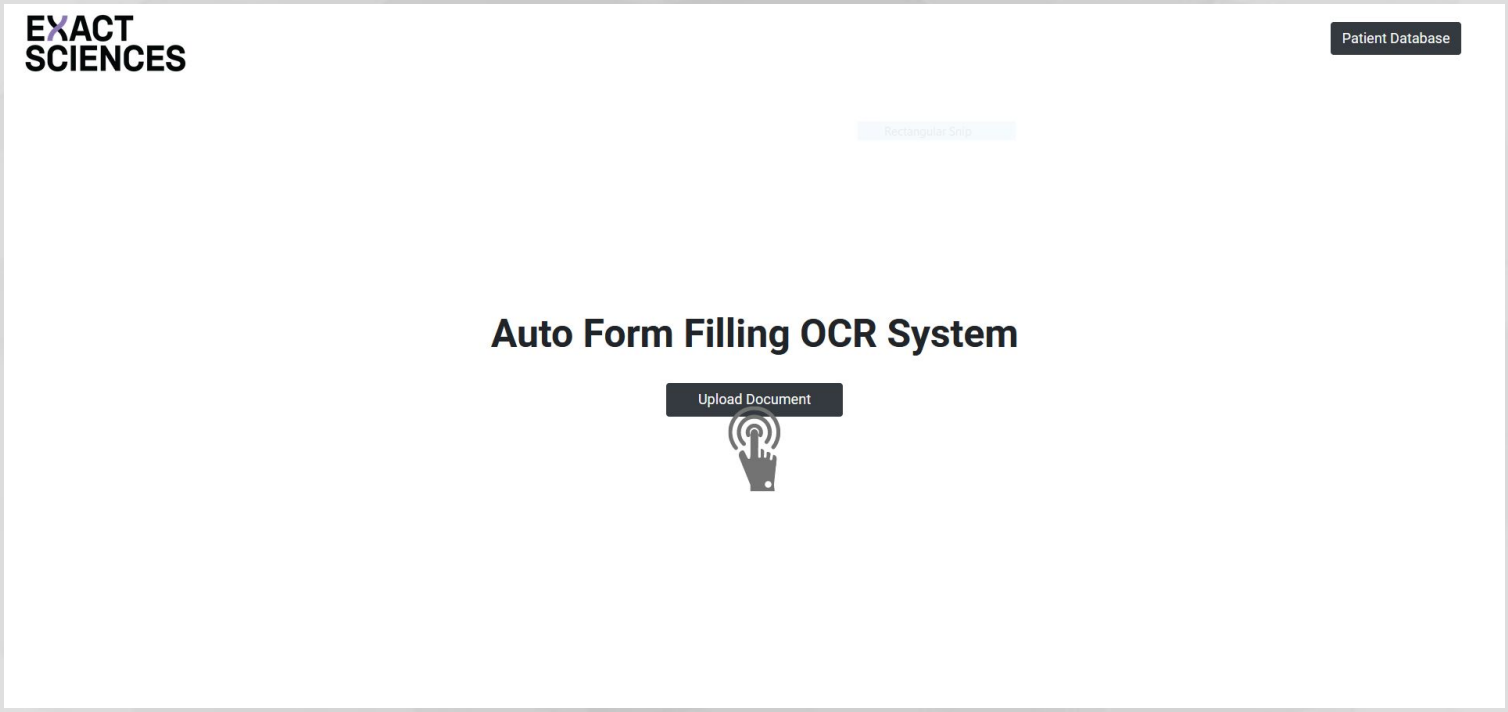
Problem of Current System

- There are technologies used for OCR solutions, but they aren’t effective and give an accurate result.
- There are different handwriting of all users. It is hard to detect it.
- The Hard form has many unnecessary things which we have to remove and get only those things which are useful.
- The Time required by the recognizer should be reduced.
- It provide around 30% accuracy to OCR form filling.

Our Approach For Problem

- Perform Google Cloud Vision API to detect the text from the Form.
- Run Algorithm API on many different size forms filled by us.
- Detect Region of Interest(ROI) using contour detection.
- Easy web portal to upload a PDF document.
- Edit, view, delete and download documents. (CRUD Application)

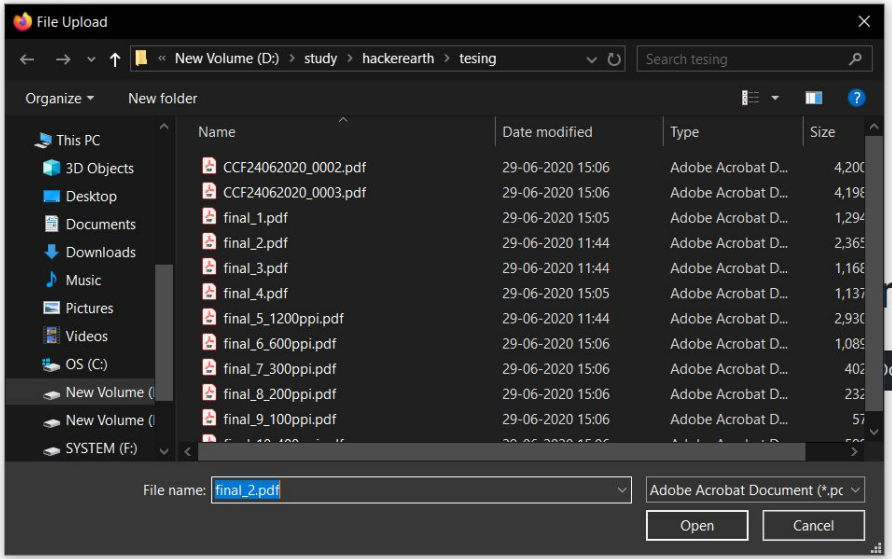
How we did it...



Home Page snapshot

EXACT SCIENCES

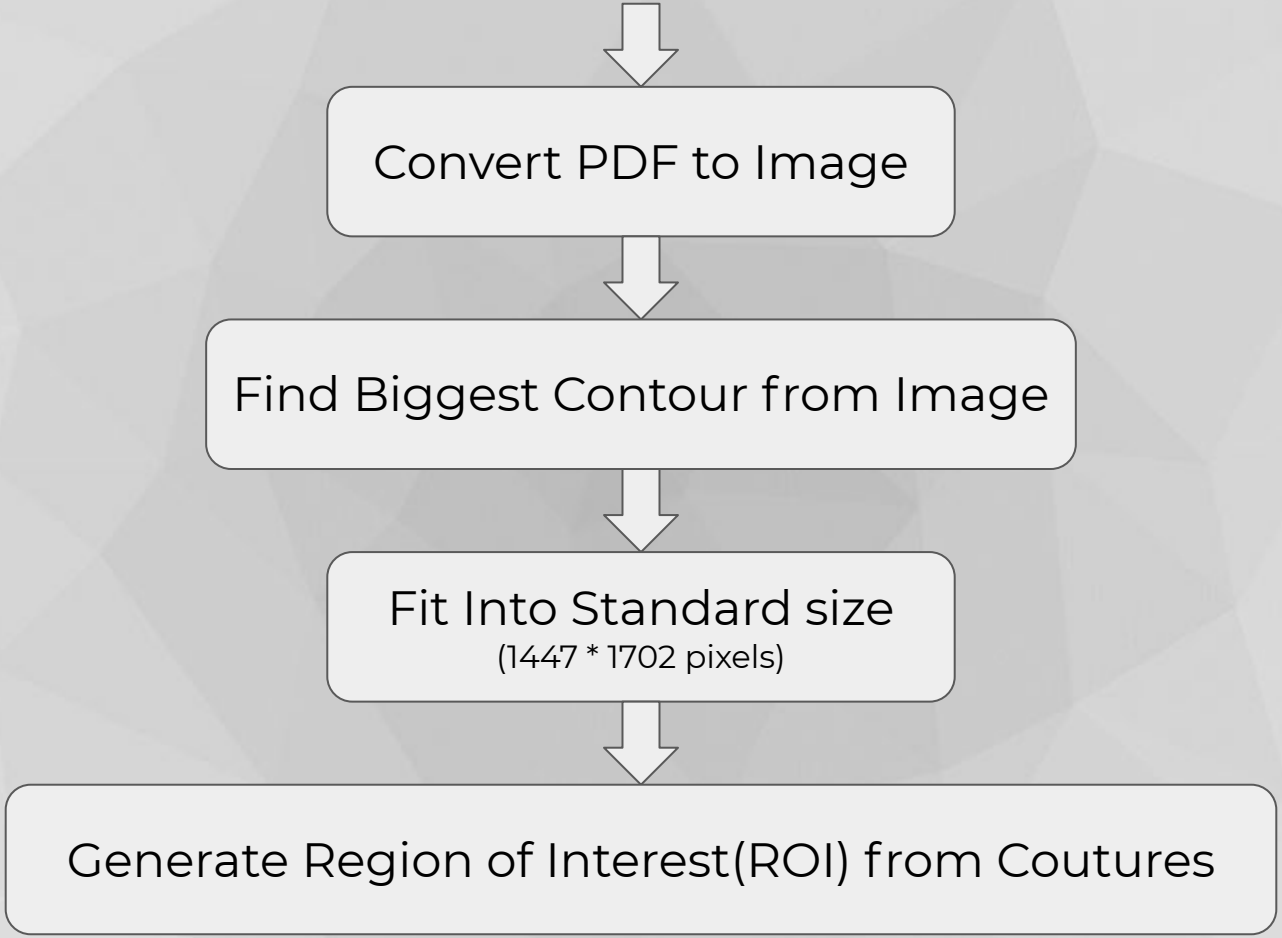
Patient Database

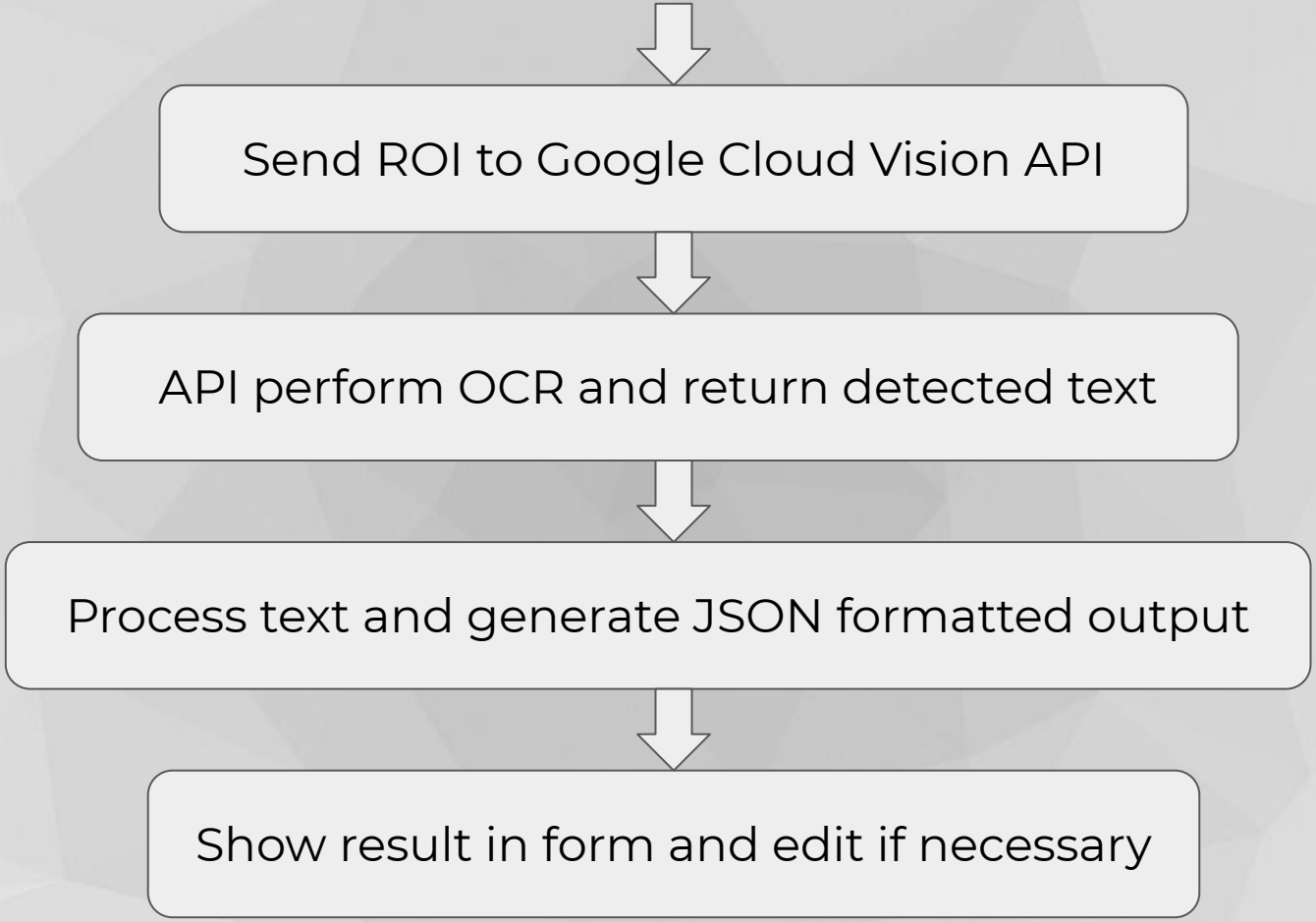


ng OCR System

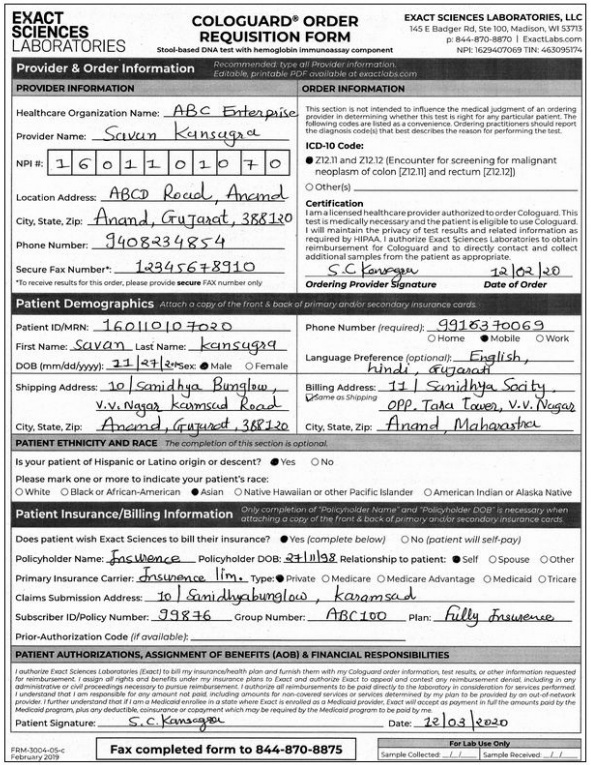
ocument

Select pdf file on that the OCR will run


Convert PDF to Image
Find Biggest Contour from Image
Fit Into Standard size
(1447 * 1702 pixels)
Generate Region of Interest(ROI) from Coutures


Send ROI to Google Cloud Vision API
API perform OCR and return detected text
Process text and generate JSON formatted output
Show result in form and edit if necessary

Hand written and Digital Form Comparison



EXACT SCIENCES LABORATORIES

COLOGUARD® ORDER REQUISITION FORM

Stool-based DNA test with hemoglobin immunoassay component

EXACT SCIENCES LABORATORIES, LLC
145 E Badger Rd, Ste 100, Madison, WI 53713
p: 844-870-8870 | ExactLabs.com
NPI: 1629407069 TIN: 463095174

Provider & Order Information *Recommended: type all Provider information. Editable, printable PDF available at exactlabs.com*

PROVIDER INFORMATION

Healthcare Organization Name: ABC Enterprise
Provider Name: Savan Kunsugea
NPI #: 1 6 0 1 1 0 1 0 7 0
Location Address: ABCD Road, Anand
City, State, Zip: Anand, Gujarat, 388120
Phone Number: 9408234854
Secure Fax Number*: 12345678910
*To receive results for this order, please provide **secure** FAX number only

ORDER INFORMATION

This section is not intended to influence the medical judgment of an ordering provider in determining whether this test is right for any particular patient. The following codes are listed as a convenience. Ordering practitioners should report the diagnosis code(s) that best describes the reason for performing the test.

ICD-10 Code:
● Z12.11 and Z12.12 (Encounter for screening for malignant neoplasm of colon [Z12.11] and rectum [Z12.12])
○ Other(s) _____

Certification
I am a licensed healthcare provider authorized to order Cologuard. This test is medically necessary and the patient is eligible to use Cologuard. I will maintain the privacy of test results and related information as required by HIPAA. I authorize Exact Sciences Laboratories to obtain reimbursement for Cologuard and to directly contact and collect additional samples from the patient as appropriate.

S.C Kunsugea — *Ordering Provider Signature*
12/02/20 — *Date of Order*

Patient Demographics *Attach a copy of the front & back of primary and/or secondary insurance cards.*

Patient ID/MRN: 160110107020
First Name: Savan Last Name: Kunsugea
DOB (mm/dd/yyyy): 11/27/2000 Sex: ● Male ○ Female
Shipping Address: 10 | Sanidhya Bunglow, V.V. Nagar Kamsad Road
City, State, Zip: Anand, Gujarat, 388120

Phone Number *(required)*: 9916370069 ○ Home ● Mobile ○ Work
Language Preference *(optional)*: English, Hindi, Gujarati
Billing Address: 11 | Sanidhya Society, OPP. Tata Tower, V.V. Nagar
☑ Same as Shipping
City, State, Zip: Anand, Maharashtra

PATIENT ETHNICITY AND RACE *The completion of this section is optional.*

Is your patient of Hispanic or Latino origin or descent? ● Yes ○ No
Please mark one or more to indicate your patient's race:
○ White ○ Black or African-American ● Asian ○ Native Hawaiian or other Pacific Islander ○ American Indian or Alaska Native

Patient Insurance/Billing Information *Only completion of "Policyholder Name" and "Policyholder DOB" is necessary when attaching a copy of the front & back of primary and/or secondary insurance cards.*

Does patient wish Exact Sciences to bill their insurance? ● Yes *(complete below)* ○ No *(patient will self-pay)*
Policyholder Name: Insurence Policyholder DOB: 27/11/98 Relationship to patient: ● Self ○ Spouse ○ Other
Primary Insurance Carrier: Insurence Ltm. Type: ● Private ○ Medicare ○ Medicare Advantage ○ Medicaid ○ Tricare
Claims Submission Address: 10 | Sanidhyabunglow, Karamsad
Subscriber ID/Policy Number: 99876 Group Number: ABC100 Plan: Fully Insurence
Prior-Authorization Code *(if available)*: _____

PATIENT AUTHORIZATIONS, ASSIGNMENT OF BENEFITS (AOB) & FINANCIAL RESPONSIBILITIES

I authorize Exact Sciences Laboratories (Exact) to bill my insurance/health plan and furnish them with my Cologuard order information, test results, or other information requested for reimbursement. I assign all rights and benefits under my insurance plans to Exact and authorize Exact to appeal and contest any reimbursement denial, including in any administrative or civil proceedings necessary to pursue reimbursement. I authorize all reimbursements to be paid directly to the laboratory in consideration for services performed. I understand that I am responsible for any amount not paid, including amounts for non-covered services or services determined by my plan to be provided by an out-of-network provider. I further understand that if I am a Medicaid enrollee in a state where Exact is enrolled as a Medicaid provider, Exact will accept as payment in full the amounts paid by the Medicaid program, plus any deductible, coinsurance or copayment which may be required by the Medicaid program to be paid by me.

Patient Signature: S.C Kunsugea Date: 12/03/2020

FRM-3004-05-c February 2019

Fax completed form to 844-870-8875

For Lab Use Only
Sample Collected: __/__/__ Sample Received: __/__/__

EXACT SCIENCES

COLOGUARD ORDER REQUISITION FORM

cologuard

Provider & Order Information

Provider Information

Healthcare Organization Name : ABC Enterprise

Provider Name : Savan Kunsugea

NPI # : 070I0II09I'

Location Address Name : ABCD Road, Anand

City, State, Zip : Anand, Gujarat, 388120

Phone Number : 9408234854

Secure Fax Number : 12345678910der,p1ea5epr0v1de5e0ureFAXnumber0n1y

Order Information

☐ Please mark right only if patient signature is on original hard form.(Optional)

NEXT

Your PDF has one page only.

This page is optional.

EXACT SCIENCES

COLOGUARD PROCEDURE INFORMATION

cologuard

Information from Cologuard Provider

Cologuard Order Number :	
Date Received by ES Labs :	
Health Organization Name :	ABC Enterprise
Provider Name :	Savan Kunsugea
Provider NPI # :	0701011091
ICD-10 Code :	ion
Patient Name :	Hdbood11 grMaloomalo
Patient Date of Birth :	
Patient Sex :	Female
Patient Phone Number :	H0meM0b11e0W0rk

☐ Please mark right only if all data extracted are as same as hard copy for both form.

SUBMIT

Till Now It tacks Average 30-35 Seconds for processing.

CRUD Application & Download feature

EXACT SCIENCES

Patient Details

#	Patient ID	Patient Name	Provider Name	Provider NPI #	Last Update	Actions
1	160110107021	keyur khant	keyur khant	1601170211	29/06/2020 12:10:03	
2	160110107020	Savan kansugra	Savan kansagra	1601101070	29/06/2020 19:42:46	

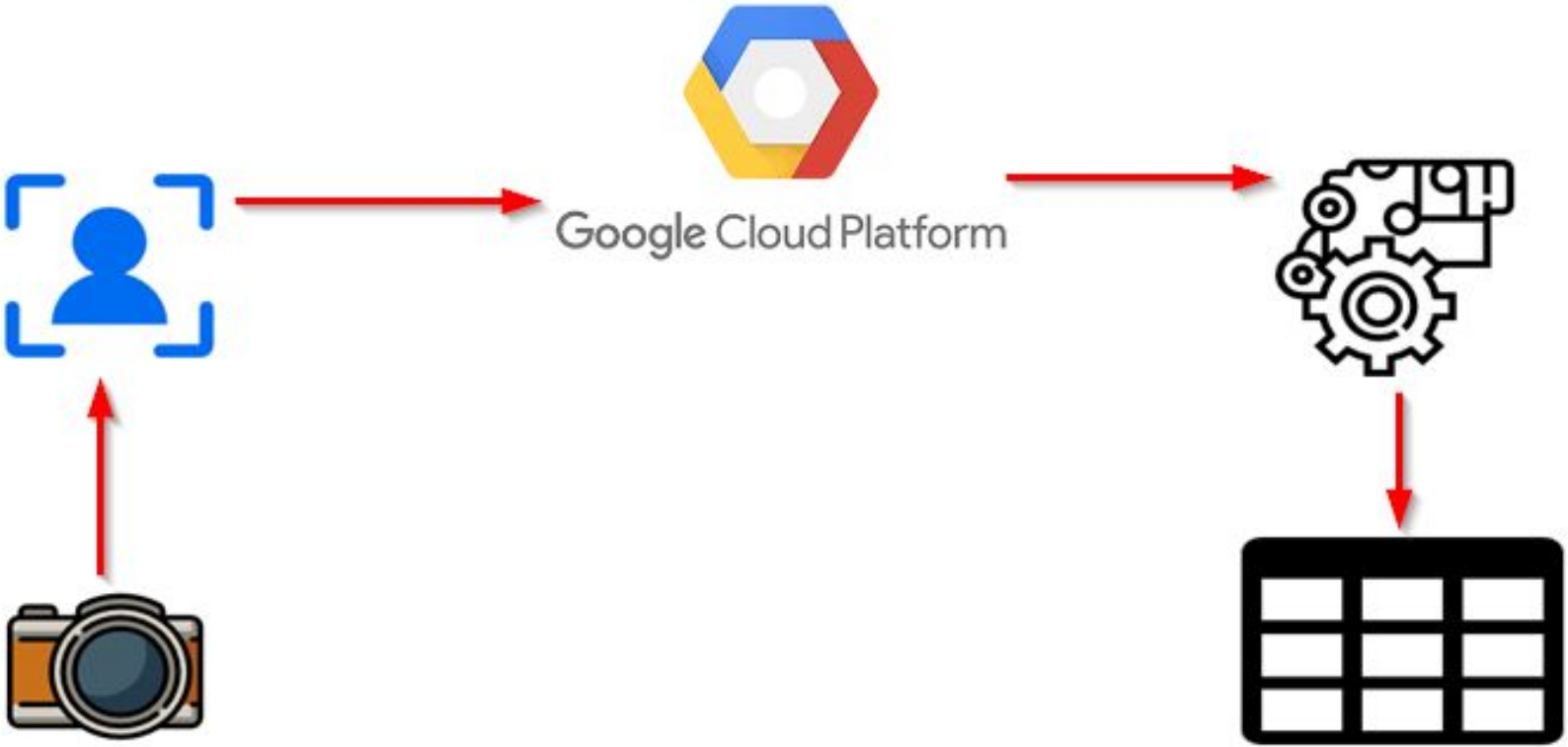
All Patient Details with view, edit, delete and download Options.

Tools and Technologies used

Optical Character Recognition(OCR) Using Google Vision API

Google Vision API

- Document Text Detection method to detect handwritten text.
- Easy to use with good accuracy
- Optimized for dense areas of text in an image


Google Cloud Platform

OpenCV

Python

- OpenCV library used for image processing like Contour Generation and Hough Line Detection.
- Python used for text processing.

Database

MongoDB Database

- Open Source NoSQL database
- Store Key-Value Pair JSON Object
- High Performance

Web Application

Flask(Python) Framework

HTML,CSS, JavaScript

Back-end Development

Front-end Development

What is different?

- It gives effective recognition results of both text and handwritten characters.
- User friendly & responsive GUI through which any individuals can upload patient information on the database.
- Minimum dependency on staff for manual entries.
- The System which gives many effective & useful features like -
  - Upload or scan the document easily
  - Automated form fillup
  - Preview of original form for cross-checking.
  - CRUD operations.
  - Store information on the database

Cost of Implementation

- Computer With Internet Connection
- Scanner
- Google Cloud Vision API

Feature	Price per 1000 units		
	First 1000 units/month	Units 1001 - 5,000,000 / month	Units 5,000,001 - 20,000,000 / month
Label Detection	Free	$1.50	$1.00
Text Detection	Free	$1.50	$0.60
Document Text Detection	Free	$1.50	$0.60

Google cloud vision API price table. Reference: https://cloud.google.com/vision/pricing

If you perform **30000 API calls per month** then it costs **30*1.50= $45 per month.**

Result of 4 different forms

No. of Pages in Document	Average Fully Accurate Fields	Average Accuracy	Average Time Required
One Page(No Information Needed Form)	33.5 out of 38	88.15%	29.8 sec
Two Pages(Both Form)	36 out of 40	90%	38 sec

Github Repository

https://github.com/keyurkhant/OCR-Auto-Form

Thank You