

Extracting Data From Fax Formatted Form Using OCR

Objective:

In the era of digitalization, most of the things are done on digital devices and platforms. But many areas where there is still manual work is used like form fillings of any individuals. There are technologies like Optical Character Recognition to convert image format into text but lack effectiveness it doesn't use. Here we try our best to make an effective product which used to take standard fax formatted form of patients and store it in a database for future use.

How to run application:

1. Extract zip file or clone a given github repository into your system.

<https://github.com/keyurkhant/OCR-Auto-Form.git>

2. Install virtual environment python package & create a virtual environment named **VisionAPI** into **OCR-Auto-Form** and activate the virtual environment.

3. Install **VisionAPI/requirements.txt** into virtual environment using following command:

```
pip install -r requirements.txt
```

4. For **Windows** and **Mac OS**, install **Poppler** which is given into a zip file.

Note: There are many dependencies which can't be included into requirements.txt file or platform dependant. So install it as required.

5. Run **final.py** flask application.

6. URI is stored in **final.py** file & **MongoDB** Compass access:

Username: kpkhant

Password: keyurkhant123

7. For **Google Vision API**, you can use default configuration of our cloud account. But future use install and configure API from **Google Cloud Console**.

API Credential JSON File : VisionAPI/VisionModule/ServiceAccountToken.json

This VisionAPI gives 1000 API calls free each month and charges \$1.5 after for each 1000 calls.

Default configuration(Our Account) has \$50 credit. Using it you can process around 950 forms.

For Configuration of your own Cloud VisionAPI into the system.

Reference: <https://cloud.google.com/vision/docs/setup>

Technologies required:

1. Google Cloud VisionAPI
2. MongoDB Database
3. Computer Vision (OpenCV library)
4. Flask (Python Backend Framework)
5. HTML, CSS, JavaScript, jQuery

Solution:

Our main objective is to **detect specific text (Handwritten)** from the image or pdf file. So our approach for that problem is as follow:

Step 1: Take a PDF file as input and convert it into numpy images.

Step 2: Then take the row image(Original), find contours in that image and take the largest one.

Contour is one **line/rectangle detection** algorithm in **Computer Vision**. We used the **OpenCV** library and found contours for both images.

There are many contours but we required which has largest area because it shows from area and its result is as follow for both images:

Provider & Order Information	
PROVIDER INFORMATION	ORDER INFORMATION
Healthcare Organization Name: <u>Exact Science</u>	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.
Provider Name: <u>Keyur Khant</u>	ICD-10 Code:
NPI #: <u>1234567891</u>	● Z12.11 and Z12.12 (Encounter for screening for malignant neoplasm of colon [Z12.11] and rectum [Z12.12])
Location Address: <u>123, Gayatri Soc.</u>	○ Other(s)
City, State, Zip: <u>Surat, Gujarat</u>	Certification
Phone Number: <u>9016243435</u>	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.
Secure Fax Number: <u>129660</u>	Ordering Provider Signature <u>Keyur Khant</u> Date of Order <u>12/02/2020</u>
Patient Demographics	Language Preference
Patient ID/MRN: <u>959595</u>	Phone Number (required): <u>9016243439</u>
First Name: <u>Khushal</u> Last Name: <u>Patel</u>	● Home ○ Mobile ○ Work
DOB (mm/dd/yyyy): <u>12/02/1978</u> Sex: ● Male ○ Female	Language Preference (optional): <u>English</u>
Shipping Address: <u>31, Gayatri Society, Surat</u>	Billing Address: <u>Gayatri Society</u>
City, State, Zip: <u>Surat, 395006</u>	City, State, Zip:
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	
Does patient wish Exact Sciences to bill their insurance? ● Yes (complete below) ○ No (patient will self-pay)	
Policyholder Name: <u>Khushal</u> Policyholder DOB: <u>12/02/1978</u> Relationship to patient: ● Self ○ Spouse ○ Other	
Primary Insurance Carrier: <u>XYZ</u> Type: ○ Private ● Medicare ○ Medicare Advantage ○ Medicaid ○ Other	
Claims Submission Address: <u>971, XYZ P Society, Delhi</u>	
Subscriber ID/Policy Number: <u>909091</u> Group Number: <u>35A</u> Plan: <u>13961AB</u>	
Prior Authorization Code (if available): <u>616121913</u>	
PATIENT AUTHORIZATIONS, ASSIGNMENT OF BENEFITS (AOB) & FINANCIAL RESPONSIBILITIES	
I authorize Exact Sciences Laboratories (Exact) to bill my insurance/health plan and to furnish them with my Cologuard order information, test results, or other information requested for reimbursement. I assign all rights and interest in my insurance plan to Exact and authorize Exact to assign and collect any reimbursement amount, including any administrative or civil proceedings necessary to pursue reimbursement. I authorize all reimbursements to be paid directly to the laboratory in consideration for service performed. I understand that I am releasing my right to 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, copayment or co-insurance which may be required by the Medicaid program to be paid by me.	
Patient Signature: <u>Keyur Khant</u>	Date: <u>12/02/2020</u>

Cologuard Order Number:	
Date Received by ES Labs:	<u>6/09/2019</u>
Health Organization Name:	<u>123 Healthcare</u>
Provider Name:	<u>Rob Pizza, MD</u>
Provider NPI:	<u>1134225618</u>
ICD-10 Codes Z12.11 and Z12.12: (Encounter for screening for malignant neoplasm of colon [Z12.11] and rectum [Z12.12]) The above codes are listed as a convenience. Ordering practitioners should report the diagnosis code(s) that best describes the reason for performing the test, regardless of whether the code is listed above or not.	<u>Z12.11</u> <u>Z12.12</u>
Patient Name:	<u>Allie <Last Name></u>
Patient Date of Birth:	<u>6/9/1954</u>
Patient Sex:	<u>Female</u>
Patient Phone Number:	<u>608-555-1003</u>
Patient Shipping Address:	<u>1440 Monroe St Madison WI 53711</u>
Please Confirm Secure Fax #: For Results and Patient Information	<u>608-867-5309</u>
Healthcare Provider Signature: Please Sign this field if blank. We must have a valid Provider Signature to proceed.	<u>Yes</u>
Insurance Type: (Medicare, Medicare Advantage, Medicaid, Insurance, Self-Pay)	<u>Medicare Advantage</u>
Insurance Carrier Name: (Example: Blue Cross, Aetna) Please add the Claims address or fax a copy of the insurance card	<u>Turner & Hooch</u>
Subscriber ID:	<u>8675309</u>
Group Number:	
Policy Owner/Holder Name:	<u>Allie <Last Name></u>
Policy Owner/Holder Date of Birth:	

Figure 1 Largest Area Contour for both forms

Step 3: Generate **Region of Interest (ROI)** for required 35 fields for both images combined.

It show results as follow:

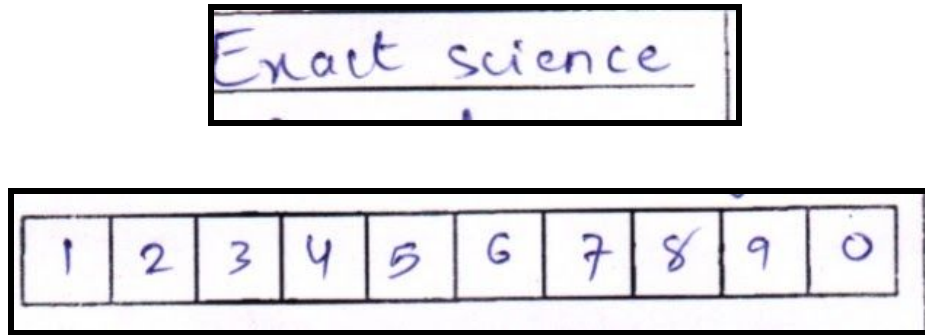


Figure 2 Region of Interest of Provider and NPI#

Step 4: Apply **Google Vision API Document Text Detection** method to detect handwritten text in generated ROIs. Generation of ROIs depends on the quality of scanned images and image PPI(Pixels per Inch). So a better scanned image gives an accurate result.

Step 5: Generate **JSON** format file and give it to **Flask**.

Step 6: For Web Application, **Flask(Python)** is used for backend and simple **Javascript & jQuery** for frontend. **MongoDB** for data storage and **CRUD** application.

In **CRUD** Application, you can **view**, **edit** and **delete** patient entries.

Step 7: **Download** digital filled cologuard form using **pdfkit** python library.

Result:

Final accuracy mainly depends on the quality **of handwriting** and **proper scanned document**. It gives an error if the main contour can't be found by the contour detector and it is possible when the scanned document is very blurry or some part of the document can't be found.

System is fully automated. You have to just upload a PDF file and it gives a resulting filled form.

Here, we test automation system with numbers of scenario with 4 filled form and we conclude following result:

Tabel 1 Final Result of 4 filled 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

Time required depends on internet speed.

Image vs JSON Format Output:

Provider & Order Information	
PROVIDER INFORMATION	ORDER INFORMATION
Healthcare Organization Name: <u>Exact Science</u>	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.
Provider Name: <u>Keyur Khant</u>	ICD-10 Code:
NPI #: <u>1234567891</u>	● Z12.11 and Z12.12 [Encounter for screening for malignant neoplasm of colon [Z12.11] and rectum [Z12.12]]
Location Address: <u>123, Gayatri Soc.</u>	○ Other(s):
City, State, Zip: <u>Surat, Gujarat</u>	Certification
Phone Number: <u>9016243435</u>	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.
Secure Fax Number: <u>129660</u>	<u>Keyur</u> <u>12/02/2020</u>
*To receive results for this order, please provide secure FAX number only.	
Patient Demographics	
Patient ID/MRN: <u>959595</u>	Phone Number (required): <u>9016243439</u>
First Name: <u>Khushal</u> Last Name: <u>Patel</u>	● Home ○ Mobile ○ Work
DOB (mm/dd/yyyy): <u>12/02/1998</u> Sex: ● Male ○ Female	Language Preference (optional): <u>English</u>
Shipping Address: <u>31, Gayatri Society</u>	Billing Address: <u>Gujarat Society</u>
<u>Surat</u>	✓ Same as Shipping
City, State, Zip: <u>Surat, 395006</u>	City, State, Zip:
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	
Does patient wish Exact Sciences to bill their insurance? ● Yes (complete below) ○ No (patient will self-pay)	
Policyholder Name: <u>Khushal</u> Policyholder DOB: <u>12/02/1998</u> Relationship to patient: ● Self ○ Spouse ○ Other	
Primary Insurance Carrier: <u>XYZ</u> Type: ○ Private ● Medicare ○ Medicare Advantage ○ Medicaid ○ Tricare	
Claims Submission Address: <u>991, XYZP Society, Delhi</u>	
Subscriber ID/Policy Number: <u>909091</u> Group Number: <u>35A</u> Plan: <u>1B96 LAB</u>	
Prior-Authorization Code (if available): <u>616121913</u>	
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: <u>Keyur</u> Date: <u>20/06/2020</u>	

Figure 3 Image Input

```
"form1": {
  "hco_name1": "Exact Science ",
  "provider_name1": "Keyur Khant ",
  "pr_npil": "1234567891",
  "pr_address1": "1231 Gayatai Soc",
  "pr_city1": "Surat, Gujarat ",
  "pr_phone1": "9016243435",
  "pr_fax1": "129660000000",
  "icd_codel": "default",
  "icd_other1": "",
  "order_date1": "12/02/2020",
  "pt_id1": "959595",
  "pt_fname1": "Khushal",
  "pt_lname1": "Patel",
  "pt_dob1": "12/02/1998",
  "pt_sex1": "Male",
  "pt_phone1": "9016243439",
  "pt_phonetype1": "Home",
  "pt_lang1": "English",
  "pt_address1": "31, Gayatri Surat. Society ",
  "pt_scity1": "Susat, 395006 ",
  "pt_baddress1": "Myarat son nety ",
  "pt_bcity1": "",
  "pt_latino1": "Yes",
  "pt_race1": ["American Indian or Alaska Native"],
  "pt_bill1": "Yes",
  "poly_name1": "Khushal",
  "poly_dob1": "12/02",
  "pt_relation1": "Self",
  "insurance_carrier1": "Zhx",
  "insurance_type1": "Medicare",
  "claim_address1": "991, XYZP fociety, Delhi 22 ",
  "sub_id1": "909091",
  "group_number1": "35A",
  "plan1": "1B96 LAB ",
  "auth_codel": "616121913",
  "final_date1": "20/06/2020",
```

Figure 4 Result as JSON Format

Cost of Implementation & Feasibility

- All the tools and technologies which are used to develop this system are free of cost except Google Vision API.
- Google vision API gives free access for some API calls, after it charges minimally which is affordable for any organization, specially hospitals.
- We can make our own OCR system but with the lack of datasets, there are high chances to get errors in recognized results.
- So, Google Vision API is the best of all available softwares.
- All revivify requires is a good internet connection, which makes it affordable for everyone.

Team TheRenaissance

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