# AUTOMATED DATA ENTRY USING OCR

<sup>1</sup>Sangita Chaudhari, <sup>2</sup>Pooja Solanki, <sup>3</sup>Tirth Trivedi, <sup>4</sup>Pratik Wadekar

<sup>1</sup>Assistant Professor, <sup>2</sup>Student, <sup>3</sup>Student, <sup>4</sup>Student <sup>1</sup>BE Computer Engineering, <sup>1</sup> Vidyavardhini's College of Engineering and Technology, Vasai West, India

Abstract: Forms serve as the most widely used asset to collect information. Different institutions and organizations use forms to collect information from the other end. In earlier times, paper forms remained the elementary method to collect information. With the passage of time, the mode of collecting information evolved gradually. Currently the data from the users are collected online with the help of various applications and online tools. But still a huge chunk of information is still offline in the paper forms and needs to be transferred online for various purposes. With the idea of storing data in digital form, manually entering the data into the database was the only prevalent method and it consumed a lot of human effort and time. The intention was to transfer the data online with minimum manual work and therefore Optical Character Recognition (OCR) came into the picture. The central idea is that the data from the paper forms which needs to be stored online will be scanned with a device and the digital data from the forms will get converted to machine suitable format (strings) with the help of OCR and image processing techniques. The purpose behind writing this paper is to present the idea of how data entry process can be made easy with help of automated techniques.

# Keywords: OCR, data entry, fields, database

# 1. INTRODUCTION

Filling out forms is one of the oldest and widely used methods for collecting information in different fields from the applicant. Automating the scanning process of large volume of office data such as cheques, aadhar card forms, driving license forms, new account opening forms, can escalate the office productivity as well as reduce time consumption. With the recent advances in technology, the manual filling of data is widely replaced by computerized data. Almost all the forms and data are submitted online. Also there is a deep requirement that the information collected offline using form filling should be available online for faster access in near future. Data available online can also be easily manipulated. Automation is basically demanded in places such as in income tax offices, banks, post office, municipal department, colleges, university, where large amount of data is to be manipulated. This problem is very recent as there is a rapid emergence of data collection offline. Many researchers are working over this issue and have developed numerous algorithms. Forms that estimate form data and handwritten data automatically are usually more error prone. It is always beneficial to first convey to the system about the form from which data is to be extracted. That is why, handwritten data extraction system which is form specific is more accurate and will extract data with lesser errors. Moreover, it has been observed that in offices the forms that are being distributed are static, which means the fields do not change rapidly over time. So, it is again a good idea to go for a handwritten data extraction system which is form specific.

In majority of the offices data entry is still offline. They are collected over sheet of paper and then typed back into computer manually. Automating this process with the help of computer vision may include several steps. One of the major tasks is extracting the handwritten data from the application form. The extracted data can be used for many purposes for example archiving and documenting. The extracted data can also be given to optical character recognition engine to convert it to corresponding Unicode number. This will help organize data and may improvise data processing.

Recognizing relative locations of information within form is another important process. Some papers suggest that by recognizing lines using Histogram techniques to estimate the location of data, but this may fail if the lines are hiding behind a content occupying several lines at front. This process can be improved if the form format is known. Template matching can be used for analyzing the relative position of data fields and then estimating the location of handwritten data, this will improve the probability of finding required data accurately, reducing false positive results.

Another commonly found element in an application form is straight lines. So their extraction is very important. Straight lines are often found around data to be extracted. Straight line detection problem is often found when we are designing a system where data to be extracted is in a completely unknown form format or there is a very huge skewness or rotation is involved. This problem can be reduced if form format is already known and the user is instructed to perform scanning operation in a controlled environment. This kind of setup will not only improve the accuracy but also increase the flexibility to involve not only the handwritten text data but also signatures, fingerprints, color photographs and so on to be entered into database.

Selection of appropriate feature recognition method is the most important aspect for data extraction from form images. Several methods for feature recognition have been proposed till date.

# 2. STUDIED SYSTEMS

 $2.1\,$  A handwritten data extraction system based on common patterns like lines bounding the filled data

Here the system rigorously searches for straight lines in both horizontal and vertical direction and then decides area bounding handwritten data. This approach has limitation over the type of form that it can recognize. The entered data must be inside bounding rectangle, forms having straight lines or no lines can be difficult for this system to recognize. Though this method has good flexibility over handling scaling and skew, but it may consume more time as it requires running a CPU intensive operation like line detection using Hough line transform.

#### © 2019 JETIR April 2019, Volume 6, Issue 4

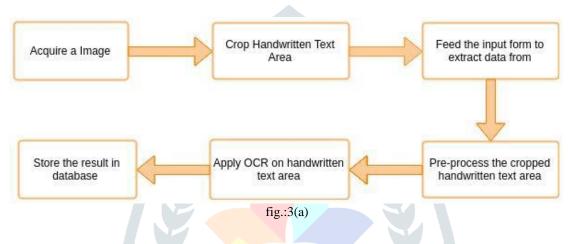
### 2.2 Extraction of handwritten data based on color of ink used to fill data

For different set of field a variety of colors are used to fill them. During processing the system recognizes the color of filled in data and extract them accordingly. Handwritten data which collides with base form foreground is gone through a series of processing for the reconstruction of extracted handwritten data. This method is easy to implement but there is a serious drawback, it requires the person filling the form to use different ink colors for distinct fields, which can be very tedious and may consume more time for users. After all these drawbacks like limiting users, this approach have very good efficiency as all the data that needs to be extracted have distinct features. Making it easy to recognize and separate. Also this method has good tolerance towards skew and nonuniform scaling.

# 2.3 Using geometric transformations to separate data.

A proposed system in which the system requires correction of geometric transformations within form, maintaining a standard dimension, it then subtracts the complete form template with the corrected template. The correction of input form fixes various kinds of transformations like rotation, non-uniform scaling, left skew, right skew. Proposed system makes use of histogram and various other techniques for correcting transformations. This method has a strong drawback that it is subtracting the complete form template to the input form. This will in most of the cases may leave unwanted data around the handwritten text, or it may also include form field data, if transformations are not resolved correctly.

# 3. BACKGROUND



#### 3.1 ACQUIRING A TEMPLATE IMAGE:

- The sample form can be taken from commonly used input devices like a scanner, digital camera or any other digital input device.

- For this proposed system it is advised to use scanner since it produces less commonly found image transformations, for example, shear, rotation, scaling and reflection.

- After acquiring a sample form, template images for each of the data fields are extracted manually.

- After extracting a template image we then feed in the coordinates of the handwritten text area

#### 3.2 CROP HANDWRITTEN AREA:

- The required handwritten area is cropped so that time is not wasted on preprocessing the part which is not required.

- This stage saves the further time in conversion of extracted data.

# 3.3 FEED THE INPUT FORM TO EXTRACT DATA FROM:

- Once the template form is scanned and coordinates are determined, the actual form whose data is to be inputted in the system is scanned.

# 3.4 PREPROCESSING:

- The input form scanned from input device is raw.

- Before actually processing the input image, it is necessary to process the image in order to remove the different types of noises present in the image.

# © 2019 JETIR April 2019, Volume 6, Issue 4

- Steps involved in preprocessing are –

#### i) Grey Conversion

A RGB image is heavily dynamic with even a slight change in the environment.

With grey channel there is a slightly less impact on image from the surrounding changes. Pixels can be represented with values ranging between 0 -255

Pixelvalue = 0.3R + 0.6G + 0.1B

#### ii) Binarization

The grey image is converted into binary image having only two pixel values i.e. 0(black) and 255(white). It separates foreground image with background image.

#### iii) Noise Removal

After binarization, some noise might be present in it which may contribute to false recognition and unnecessary computations. Involves 2 steps:

#### 1. Dilation

-In dilation the brighter regions i.e. pixels having value of 255 are expanded leaving behind the darker spots with lower diameter to shrink and disappear as shown in fig 3(b).

- Dilation is nothing but performing UNION operation on the input image and structuring 16 element  $(A \oplus B)$ .

- Performing only dilation will also result in reduction of number of black pixels in the text foreground, which might reduce the recognition accuracy of extracted field template in the first step with input for due to reduction in number of useful foreground pixels. So to overcome this problem we will perform erosion over the image obtained after dilation.

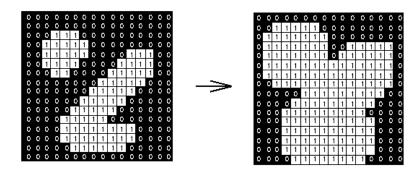


fig.:3(b) dilation

#### 2. Erosion

- Erosion is the process which just opposite to dilation. It expands the black region which in turn reduces the white region as shown in fig 3(c).

- The gaps in the letters of the text are produced because of dilation will fill up as well. This will recover most of the damage done by dilation

- Erosion is mostly "AND" operation between input image and structuring element

- Erosion = A  $\ominus$  B / A  $\cap$  B

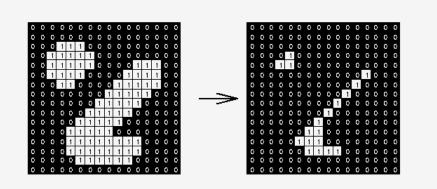


fig.:3(c) erosion

#### © 2019 JETIR April 2019, Volume 6, Issue 4

#### 3.5 APPLYING OCR:

- Once the previous image is obtained, then OCR is applied on the image to extract data from the image.

- The different OCR techniques are:

#### i. Matrix Matching

Matrix Matching converts each character into a pattern within a matrix, and then compares the pattern with an index of known characters. Its recognition is strongest on monotype and uniform single column pages.

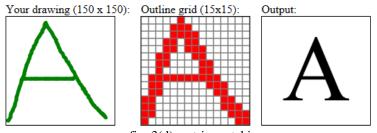
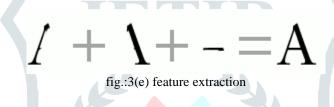


fig.:3(d) matrix matching

#### ii. Feature Extraction

This method defines each character by the presence or absence of key features, including height, width, density, loops, lines, stems and other character traits. Feature extraction is a perfect approach for OCR of magazines, laser print and high quality images.



#### iii. Neural Networks

This strategy simulates the way the human neural system works; it samples the pixels in each image and matches them to a known index of character pixel patterns. The ability to recognize the characters through abstraction is great for fixed documents and damaged text. Neural networks are ideal for specific types of problems, such as processing stock market data or finding trends in graphical patterns. In all these approaches Neural Networks are efficient than others.

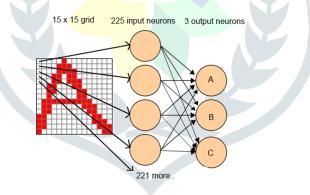
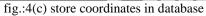


fig.:3(f) artificial neural network for single character

# 4. IMPLEMENTATION ENVIRONMENT AND RESULTS

PROCESS FORMS			
0.0	ACC CELECT & FORM THAT YO	NU WANT TO DECISTED AS A TEMPLAT	F
PLE	ASE SELECT A FORM THAT TO	DU WANT TO REGISTER AS A TEMPLAT	E
	Choose File No file chosen		
	•	UPLOAD IMAGE	
		UPLUAD IMAGE	
	~		
		GET DATA	
🕂 🔎 Type here to sea		) ڬ 💷 🛄 🗮 😐 🥥 🗾 🔤	🚸 🔬 💐 🔷 🖬 ¢i) ENG 10:51 PM 🖏
	fig	g.:4(a) uploading form	
G Google		Automated data entry-part 2 - Goo X +	- a -
← → C ① File   C;/I	Users/USER/Desktop/BE-Project/components/registe	er-template/register-template.html	x 🛛 🖾 😨 🕴
REGISTER TEMPLAT	E		
	300	xxxxxxxxxxxx	1000000000
		Manyelpada Virar (F	sh High School iast), Thane - 401 303.
YOU WANT TO	T A FORM THAT REGISTER AS A	र्षे प्राप्तिवर्ग/माध्य-स/मधील । दिमाया-२० द	1111/00-01/22350-22322 11110-2001 _Index_No-16:13:195
TEMPLATE		I FAVINC C	EPTIFICATE
Choose File Form1	gqi	(No change in any entry in this constituate shall be infringement of this requirement is fable to involve t	EACT IFTCATE Section III, Rule 17) e made except by the authority issuing it and any he imposition of a penalty such as that of nuslication)
		General Register No. 042	Serial No. 125
ELCROP RELDS	STOP CROPPING	E Mother's Name_ Protibha	THOUR EDINGTHONT
		3. Caste and Sub-caste_Hindu	Jidar 00000000000
Name of Pupil	•	1. Name of the Popt in far <u>Had ckear</u> 2. Mother's Name. <u>Prait bha</u> 3. Caste and Sub-case. <u>Hindu. C</u> 4. Pisco of Birk. <u>Nachardhi. Naigagaan</u> 6. Date of Birk. North and Yaar agoorging	adar 5. Nationative Tradian
Name of Pupil	•	28-10-1997 (Twent	y Eighth October
	+ RATE TO DATABASE	7. Last School attended Naharapi B. Date of admission 19-047-201	y Eighth October
		28.10-1993 (Turn Ninz 7. Last School attended Naling 8. Date of advision 19-01r-901 9. Progess Grood 9. Proges Grood	Legith Primary Scheel 2. 2. 2. 2. 2. 2. 2. 2. 2. 2.
		28-10-1993 Church     Nin z      Last School microide Nalional     Bair of semistric Nalional     Boar of semistric Nalional     Proges Grad      Donotat Grad      Donotat Grad      Donotat Grad      Donotat Grad	u - Eighth Colber Eighth Colber Ecca Nucly Seven Eghth Fundry School 0
REGISTER TEM		26.101933 (*Dera Normannessent (*Dera Last Bottor standess (*Dellang) Bottor statisticon (*Dellang) Deter statistic	y Eighth October English Frimery Scheel 2 2 2 2 2 2 2 2 2 2 2 2 2
REGISTER TEM	nare to database	26.40-193 Class Note Cast Store answer 2. Last Store answer 2. Cast Store answer 2. Date Storesson 2. Date Storesson	u - Eighth Colber Eighth Colber Ecca Nucly Seven Eghth Fundry School 0
REGISTER TEM	nare to database	26.101933 (*Dera Normannessent (*Dera Last Bottor standess (*Dellang) Bottor statisticon (*Dellang) Deter statistic	y Eighth October English Frimery Scheel 2 2 2 2 2 2 2 2 2 2 2 2 2
REGISTER TEM	nare to database	26.40-193 Class Note Cast Store answer 2. Last Store answer 2. Cast Store answer 2. Date Storesson 2. Date Storesson	y Eighth October English Frimery Scheel 2 2 2 2 2 2 2 2 2 2 2 2 2
REGISTER TEM	nare to database	26.40-193 Class Note Cast Store answer 2. Last Store answer 2. Cast Store answer 2. Date Storesson 2. Date Storesson	an Benkhn Continent Secan Minneky Geren) Egylch Promory Ghoel a ay 2013 ay 2014 ay 2015 ay 2015
REGISTER TEM	nate to batamase ch O Ht fig.:	4(b) register coordinates	an Benkhn Continent Secan Minneky Geren) Egylch Promory Ghoel a ay 2013 ay 2014 ay 2015 ay 2015
REGISTER TEM	ch O Hi fig.:	A Los 1933 ( <u>Incel</u> 2. Last Toroi answer: Nalžard 2. Last To	an Benkhn Continent Secan Minneky Geren) Egylch Promory Ghoel a ay 2013 ay 2014 ay 2015 ay 2015
REGISTER TEM	ALATE TO DATABASE	A Los 1933 ( <u>Incel</u> 2. Last Toroi answer: Nalžard 2. Last To	u Egihin October Cen Minch Sever) English Printer Scheel 2. 2. 2. 2. 2. 2. 2. 2. 2. 2.
reconstent text     P Type here to sear     Options Window Help     Options Window Help     office     of	ch O El fig:: fig:: Welcome X • th getCalection(calciumers).f. Here Connection Weichecter20) / El me the connection Weichecter20 / El me the conn	A(b) register coordinates	y Egilik Oring Englik Frimag Scheel 2. 2. 2. 2. 2. 2. 2. 2. 2. 2.
T-12 v Options Window Help control (5) ytem dmin indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania indiania in	Ch O Hi O fig.: Welcome X @ db.getCalector("customers).fi. % Welcome X @ db.getCalector("customers).fi. % Welc	4(b) register coordinates	the <u>Beschine</u> (Company and Messon) <u>Frankine</u> (Company) <u>Frankine</u> (Company) <u>Frankine</u> (Company) <u>All 2013</u> <u>All 2013</u>
T-12 v Options Window Help v Concetion (5) vytem local local local collections (1) custemes functions et	ch O Hi P fig.: weicherse * • 6.getCalettor(customers).fl. * Weicherse * • 6.getCalettor(customers).fl. * NordCorrection = veichots:270.7 • m/d * veicCorrection = veichots:270.7 • m/d * getCorrection	A(b) register coordinates	the Dicklin (Decision) Cecar Vincet, Geven) Freitch Promory School 21 2012 22 2012 24 2012 25 20 25 2
IT-12 W Options Window Help Control (1) w Control (1) w Control (1) w Control (1) cutor control (1) cutor cutor cutor cutor cutor cut	ch O Hi P fig:: Welcome X & dx.getColector(customer).fi. Welcome X & dx.getColector(fi.customer).fi. Welcome X & dx.getCol	A Constant State A Constant S	DE DE MARIE (De capacitadoremon)     CECA MINENTE GENERAL      Erentitadoremon (De capacitadoremon)     Erentitadoremon (De capacitadoremon)     Constructivo (De capacitadoremon)
IT-12 W Options Window Help Control (1) w Control (1) w Control (1) w Control (1) cutor control (1) cutor cutor cutor cutor cutor cut	ch O Et  fig.:  type: type: type:type:type:type:type:type:type:type:	A Loc All and A	Verschland Orenhammennen     Verschland Orenhammennen     Seren Ninnerker     Se
IT-12 W Options Window Help Control (1) w Control (1) w Control (1) w Control (1) cutor control (1) cutor cutor cutor cutor cutor cut	ch O Et C fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.: fig.:	A Loc 1933 Cloved A Loc 1933 Cloved A Loc 1930 Cl	Dependence
IT-12 W Options Window Help Control (1) w Control (1) w Control (1) w Control (1) cutor control (1) cutor cutor cutor cutor cutor cut	ch O H G fig.: ft O batamase ft O	A Loc 1933 Cloved A Loc 1933 Cloved A Loc 4980 Notice 1 A Loc 4980 Notice 1 A Loc 4980 Notice 1 A Loc 4980 Notice 1 A Loc 4980 Notice 1 B Data Internet Intern	Dependence     Operation
IT-12 W Options Window Help Control (1) w Control (1) w Control (1) w Control (1) cutor control (1) cutor cutor cutor cutor cutor cut	ALTE TO DATAMASE ch C Let C Latance		Version of the second se
IT-12 W Options Window Help Control (1) w Control (1) w Control (1) w Control (1) cutor control (1) cutor cutor cutor cutor cutor cut	ALTE TO DATAMASE ch C Lt Fig.: ft C C Lt Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.: Fig.	#8.40-033         Charad           * Loss de monos         Nobice 100           * Discost         Control 100           * Control 100         Control 100           * Conte         Cont	Version of a set of a se
IT-12 W Options Window Help Control (1) w Control (1) w Control (1) w Control (1) cutor control (1) cutor cutor cutor cutor cutor cut	ALTE TO DATAMASE ch C Lt fig :: Weissee X & de getCalection (calciners) f.f. Fires Correction (calciners) f.f. Fi	#8.40-193         Character           # Discrete         Research	Verschlie (Neuroparative)     Verschlie (Neuroparative)     Ergelick Printwerg, School     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20     20
REGISTER TEM	ALTE TO DATAMASE ch O Lit Figs: Figs: Weiczes X & dependention (automore) fin. Weiczes X & dependention (aut		v Echila Decident ecc. Ninch, Sever) English Printing, Schoel 2. 2. 2. 2. 2. 2. 2. 2. 2. 2.
P Type here to sear P Type here to sear W Options Window Help Collection (5) System domin domin cost collections (1) functions Users collections (1) functions Users collections (1) functions total	ch O Li fig.: ch O Li fig: fig: fig: webcome × • do.getColector(customers).fi. • Webcome • 0.029 sec. Key • (0) Objectid("Scibball Style&ASOchedF1cel") • (0) do.getColector(Customers).fi. • s + • (0) do.getColector(Customers).fi. • f + •	**	v Echild Oriental econ Ninch Gaven) Ecolità Frinces Gaven) Ecolità Frinces Gaven ecologica variante actività della constructione rec. 202. variante actività della constructione variante actività della con



© 2019 JETIR April 201	19, Volume 6, Issue 4	www.jetir.org (ISSN-2349-5162)	
G Google	× 🗅 Process Forms × +	- Ø X	
$\leftrightarrow$ $\Rightarrow$ C	File   Cr/Users/USER/Desktop/BE-Project/components/process-forms/process-forms.html     Choose File   form1.avg	x 🐠 🔟 🕲 🕼 :	
	UPLOAD MAGE		
	FORM DATA		
	Caste Hindu Sider		
	MothersName Prattbha		
	Pratbha		
	Name Wadekar Pratik Disambar		
	Wadekar Praik Disahgx STORE DATA		
🖷 🔎 Тур	pe here to search O H 🔒 🥹 💷 🗉 📰 😨 🦻 🔽 fig.:4(d) predicted data		

#### 5. CONCLUSION

In this work, the objective of extracting handwritten text from a filled form and then processing the extracted text with the removal of redundant data has been met successfully. The proposed system can easily deal with various kinds of input field formats like checkboxes, data tables, dotted input, which most of the already known systems found it difficult to extract. The proposed system has been implemented successfully and is rigorously tested to achieve the goals. However, the proposed system could not deal with few issues like a major degree of skew and non-uniform scaling within the scanned form. Also, data written outside the indicated area cannot be detected.

## 6. REFERENCES

[1] Piyush Jain, "Filled-in Handwritten Data Extraction from Documents based on Geometrical Features", July 2016

- [2] N. Venkata Rao, "Optical Character Recognition Technique Algorithms", 20th January 2016. Vol.83. No.2
- [3] Vishal Chourasia, "Implementation of Optical Character Recognition Using Machine Learning", June 2018. Vol.6, Issue-6,IJSCE.