

PERFORMANCE COMPARISON OF OCR TOOLS

Dr. S.Vijayarani¹ and Ms. A.Sakila²

¹Assistant Professor, Department of Computer Science, School of Computer Science and Engineering, Bharathiar University, Coimbatore.

²M.Phil Research Scholar, Department of Computer Science, School of Computer Science and Engineering, Bharathiar University, Coimbatore.

ABSTRACT:

Optical Character Recognition (OCR) is a technique, used to convert scanned image into editable text format. Many different types of Optical Character Recognition (OCR) tools are commercially available today; it is a useful and popular method for different types of applications. OCR can predict the accurate result depends on text pre-processing and segmentation algorithms. Image quality is one of the most important factors that improve quality of recognition in performing OCR tools. Images can be processed independently (.png, .jpg, and .gif files) or in multi-page PDF documents (.pdf). The primary objective of this work is to provide the overview of various Optical Character Recognition (OCR) tools and analyses of their performance by applying the two factors of OCR tool performance i.e. accuracy and error rate.

KEYWORDS:

Optical Character Recognition (OCR), Online OCR, Free Online OCR, OCR Convert, Convert image to text.net, Free OCR, i2OCR, Free OCR to Word Convert, Google Docs.

1. INTRODUCTION

Optical Character Recognition technology recognizes the text from the images automatically. It supports different types of image formats like JPG, PNG, BMP, GIF, TIFF and multi-page PDF files. OCR involves analysis of the captured or scanned images and then translate character images into character codes, so that it can be edited, searched, stored more efficiently, displayed on-line, and used in machine processes [3] . Scanned images can easily extract that text with the help of different OCR Tools. It works with images that almost consist of text in it [1]. The output of a tool is based on the type of input image. Achieving 100% accuracy is not possible, but it is better to have something rather than nothing [1]. To improve accuracy most of the OCR tools use dictionaries, recognizing individual characters then it try to recognize entire words that exist in the selected dictionary. Sometimes it is very difficult to extract text because different font size, style, symbols and dark background. If we are using high resolution documents the OCR tools will produce best results. Many OCR tools are available as of now, but only a few of them are open source and free [2]. Normally, all the OCR tools process has five important steps. They are preprocessing, segmentation, feature extraction, classification/recognition and post processing. This is depicted in Figure 1[18].

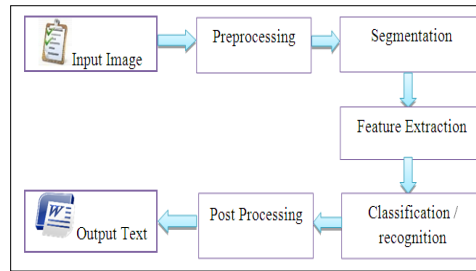


Figure 1. OCR Tools Process

Input Image

Input image is digitalized images like a scanned or captured text image. It may be of different formats, i.e. JPG, PNG, BMP, GIF, TIFF and multi-page PDF files.

Preprocessing

Preprocessing techniques are important and essential for OCR system for image handling. These techniques are used to add or remove noises from the images, maintaining the correct contrast of the image, background removal which contains any scenes or watermarks. These are applied into images which enhance the image quality. This step is essential for OCR systems [12].

Segmentation

The accuracy of OCR system mainly depends on the segmentation algorithm being used. Segmentation extracts pages, lines, words and then finally into characters from the text document images [16]. Page segmentation separates graphics from text, a line segment is a part of a line that is bounded by two distinct end points and Word segmentation is the problem of dividing a string of written language into its component words [3]. Character segmentation separates characters from others [12].

Feature Extraction

Feature Extraction stage analyzes a text segment and select a set of features that can be used to uniquely identify the text segment [18]. This stage is used to extract the most relevant information from the text image which helps to recognize the characters in the text [14].

Classification / recognition

Optical character Recognition is a most significant application. The main objective of Optical Character Recognition (OCR) is to classify the optical patterns like alphanumeric and other characters. The OCR is required when the information should be readable to both human and machine [1]. Recognition has become essential for performing classification task [13].

Post Processing

The post processing stage is used to increase recognition. The goal of post processing is to detect and correct grammatical misspellings in the OCR output text after the input image has been scanned and completely processed.

Output Text

The result of the input images is displayed in the output text.

2. OCR TOOLS COMPARISON

This paper compares eight different types of OCR tools; they are,

1. Online OCR
2. Free Online OCR
3. OCR Convert
4. Convert image to text.net
5. Free OCR
6. i2OCR
7. Free OCR to Word Convert
8. Google Docs

The main goal of this work is to compare the performance these tools for finding the best OCR tool. In order to perform the analysis, we provide an input image and this input image are processed by these OCR tools and the output produced by these tools is considered for analysis. Each OCR tools have produced different results for the same input image. The sample input image (i.e. k-means clustering algorithm) given in Figure 2 is downloaded from google images [17] and this image is used for this comparative analysis.

```

Algorithm 1 The algorithm of direct  $k$ -means clustering
  Input:  $N$  objects in the  $R$ -dimensional space, and  $M$ 
  (number of sub-cluster means)
  Output: Initial  $M$  sub-clusters that contain all  $N$  objects
  initialize  $M$  first means  $r_0^{(1)}, r_0^{(2)}, \dots, r_0^{(M)}$ 
  repeat
    for each input object  $r_i, 1 \leq i \leq N$ 
      do
        assign  $r_i$  to  $D_j$  with nearest mean  $r_0^{(j)}$ , such as
           $\varphi(r_i, r_0^{(j)}) \leq \varphi(r_i, r_0^{(u)}), 1 \leq j, u \leq M$ 
        for each sub-cluster  $D_j, 1 \leq j \leq M$ 
          do
            recalculate the mean (centroid) of objects  $r_i \in D_j$ ,
               $r_0^{(j)} = \sum_{i \in D_j} r_i / |D_j|$ , where  $|D_j|$  defines the cardinality
              of  $D_j$ 
            compute the quality function  $G$ 
          until no object has changed sub-clusters (or  $G$  does not
          change)
  
```

Figure 2 Input Image

2.1 Online OCR

OnlineOCR.net is free web-based Optical Character Recognition software (OCR) that allows, to convert scanned PDF documents (including multipage files), faxes, photographs or digital camera captured images (JPEG/JPG, BMP, PCX, PNG, GIF, ZIP file format) into editable and searchable electronic documents [4]. This tool has the capability to convert the text image in to text and this result may be displayed in different formats like Adobe PDF document, Microsoft Word document, Microsoft Excel document, RTF document and Plain Text. It supports 46 languages and has the ability to convert images to text format and its maximum input file size is 100 MB [4]. The sample input image conversion performed by Online OCR tool [4] is depicted in Figure 2.1.

```
Algorithm 1 The algorithm of direct k-means clustering N objects in the R-dimensional space, and M
(number of sub-cluster means) Output: Initial M sub-clusters that contain all N objects initialize M first
means  $r_1, r_2, \dots, r_M$  repeat for each input object  $r_i$  (1 ≤ i ≤ N) do assign  $r_i$  to  $D_j$  with nearest mean  $r_j$ ,
such as  $D_j = \{r_i \mid |r_i - r_j| < |r_i - r_k|, \forall k \neq j\}$  for each sub-cluster  $D_j$  (1 ≤ j ≤ M) do recalculate the mean (centroid) of
objects =  $\frac{1}{|D_j|} \sum_{r_i \in D_j} r_i$ , where  $|D_j|$  defines the cardinality of  $D_j$ , compute the quality function G until no object
has changed sub-clusters (or G does not change)
```

Figure 2.1 Online OCR

2.2 Free Online OCR

NewOCR.com is a free online OCR service that can analyze the text in any image file and converts the text in the image into text format. Input files supported by this tool are JPEG, JFIF, PNG, GIF, BMP, PBM, PGM, PPM and PCX. Compressed files supported by this tool are UNIX compress, bzip2, bzip and gzip. Multi page documents such as TIFF, PDF, DOCX, ODT files with images, multiple images in ZIP archive are also handled. After conversion the result has displayed in different formats, i.e. Plain text (TXT), Microsoft Word (DOC), and Adobe Acrobat (PDF). It supports 75 recognized languages and supports several font types. The advantage of Free Online OCR is, it has taken unlimited uploads. The resultant output [5] is illustrated Figure 2.2.

```
Algorithm 1 the algorithm of direct k-means clustering
Input: N objects in the R-dimensional space, and M
(number of sub-cluster means)
Output: Initial M sub-clusters that contain all N objects
initialize M first means  $r_1, r_2, \dots, r_M$ 
repeat
for each input object  $r_i$  (1 ≤ i ≤ N)
do
assign  $r_i$  to  $D_j$  with nearest mean  $r_j$ , such as
 $D_j = \{r_i \mid |r_i - r_j| < |r_i - r_k|, \forall k \neq j\}$ 
for each sub-cluster  $D_j$  (1 ≤ j ≤ M)
do
recalculate the mean (centroid) of objects  $r_i \in D_j$ 
if  $r_j = \frac{1}{|D_j|} \sum_{r_i \in D_j} r_i$ , where  $|D_j|$  defines the cardinality
of  $D_j$ ,
compute the quality function G
until no object has changed sub-clusters (or G does not
change)
```

Figure 2.2 Free Online OCR

2.3 OCR Convert

OCR Convert is a free online OCR service, which provides the facility to convert the scanned image into text. It supports JPG, PNG, BMP, GIF, TIFF and multi-page PDF files and also support low resolution images. The result may be in text format and this tool supports simultaneous uploads and able to perform conversion process of files upto 5MB (aggregated). The output text result [6] is shown in Figure 2.3.

```

Algorithm I The algorithm of direct k-means clustering
Input: N objects in the R-dimensional space. and M
(number of sub-cluster means)
Output: Initial M sub-clusters that contain all N objects
initialize M first means r$,", r$,2). ....r$;"
repeat
for each input object r,-. I €isN
do
assign r,- to D,» with nearest mean l'87, such as
<p(r,».rX) < <p(r,-.rS{'1).l € j. 11$ M
for each sub-cluster D,-. I <j<M
do
recalculate the mean (centroid) of objects r,~eD',-.
rg) = 2,-CD' r,-/|D,|. where D,- defines the cardinality
of D,-
compute the quality function G
until no object has changed sub-clusters (or G does not
change)
    
```

Figure 2.3 OCR Convert

2.4 Convert image to text.net

Convert image to text.net tool is used to convert any scanned image into editable text file with the new software JiNa OCR image to text. This software is very easy to use, just to upload an image file and click on the button it converts directly into an open word document. The output formats are Adobe PDF document, Microsoft Word document, Microsoft Excel document, Docx, HTML and Text. The output result for convert images to text.net software [7] is shown figure 2.4.

```

Algorithm 1 The algorithm of direct k-means clustering
Input: N objects in the R-dimensional space, and M
(number of sub-cluster means)
Output: Initial M sub-clusters that contain all N objects
initialize M first means 41), 42),...OF
repeat
for each input object rb I i<..A1
do
assign rito Dj with nearest mean rg), such as
49(rhe),54)(ratu),1j, u< A1
for each sub-cluster Dj, 1.<.j4M
do
recalculate the mean (centroid) of objects rie
ofDi = EiED, where |Dj| defines the cardinality
compute the quality function G
until no object has changed sub-clusters (or G does not
change)
    
```

Figure 2.4 Convert image to text.net

2.5 Free OCR

Free-OCR.com is a free online OCR (Optical Character Recognition) tool used to extract text from any image and convert these images into an editable text document. It takes a JPG, GIF,

TIFF BMP or PDF (only first page) file formats and supports 30 different languages. The only restriction of this tool is, the images must not be larger than 2MB. Output of the image [8] is illustrated in Figure 2.5

```

Algorithm 1 The algorithm of direct k-means clustering
Input: N objects in the R-dimensional space, and M
(number of sub-cluster means)
Output: Initial M sub-clusters that contain all N objects
initialize M first means  $r_1, r_2, \dots, r_M$ 
repeat
for each input object  $r_i, 1 \leq i \leq N$ 
do
assign  $r_i$  to  $D_l$  with nearest mean  $r_l$ , such as
 $\langle p(r_i, r_l) \rangle < \langle p(r_i, r_j) \rangle, \forall j \neq l, 1 \leq j \leq M$ 
for each sub-cluster  $D_l, 1 \leq l \leq M$ 
do
recalculate the mean (centroid) of objects  $r_i \in D_l$ ,
 $r_l = \sum_{r_i \in D_l} r_i / |D_l|$ , where  $|D_l|$  defines the cardinality
of  $D_l$ 
compute the quality function G
until no object has changed sub-clusters (or G does not
change)
    
```

Figure 2.5 Free OCR

2.6 i2OCR

i2OCR is a free online Optical Character Recognition (OCR) which extracts text from images and it can be edited, formatted, indexed, searched, or translated. Input image file types are TIF, JPEG, PNG, BMP, GIF, PBM, PGM and PPM. It supports 60+ Recognition Languages, major Image Formats, Multi Column Document Analysis and 100% FREE with Unlimited Uploads. The output result of the i2OCR [9] is given in Figure 2.6.

```

Algorithm 1 The algorithm of direct k-means clustering
Input: N objects in the R-dimensional space. and M
(number of sub-cluster means)
Output: Initial M sub-clusters that contain all N objects
initialize M first means  $r_1, r_2, \dots, r_M$ 
repeat
for each input object  $r_i, 1 \leq i \leq N$ 
do
assign  $r_i$  to  $D_l$  with nearest mean  $r_l$ , such as
 $\langle p(r_i, r_l) \rangle < \langle p(r_i, r_j) \rangle, \forall j \neq l, 1 \leq j \leq M$ 
for each sub-cluster  $D_l, 1 \leq l \leq M$ 
do
recalculate the mean (centroid) of objects  $r_i \in D_l$ ,
 $r_l = \sum_{r_i \in D_l} r_i / |D_l|$ , where  $|D_l|$  defines the cardinality
of  $D_l$ 
compute the quality function G
until no object has changed sub-clusters (or G does not
change)
    
```

Figure 2.6 i2OCR

2.7 Free OCR to word convert

Free OCR to Word provides a new way of translating printed text to a digital file that can be modified or edited in a word processor. The OCR to Word program works with any of the popular image files of JPG, JPEG, PSD, PNG, GIF, TIFF, BMP and scanned image files, etc. All of these file types are equally easy for Free OCR to Word and in just a few clicks, we can able to get a fully editable and searchable files in MS Word or TXT[10]. The result is shown in Figure 2.7. [10]

```

Algorithm 1 The algorithm of direct k-means clustering
Input: N objects in the R-dimensional space. and M
(number of sub-cluster means)
Output: Initial M sub-clusters that contain all N objects

initialize M first means  $r_1, r_2, \dots, r_M$ 
repeat
for each input object  $x_i, 1 \leq i \leq N$ 
do
assign  $x_i$  to  $D_j$  with nearest mean  $r_j$ , such as
 $\|x_i - r_j\| < \|x_i - r_k\|, \forall k \neq j, 1 \leq j \leq M$ 
for each sub-cluster  $D_j, 1 \leq j \leq M$ 
do
recalculate the mean (centroid) of objects  $x_i \in D_j$ 
If  $r_j = \frac{1}{|D_j|} \sum_{x_i \in D_j} x_i$ , where  $|D_j|$  defines the cardinality
of  $D_j$ 
compute the quality function  $G$ 
until no object has changed sub-clusters (or  $G$  does not
change)
    
```

Figure 2.7 Free OCR to word convert

2.8 Google Docs

Upload an image file or a scanned PDF to Google Docs, it Converts text to Google Docs format and Google Docs will automatically perform OCR on the file before saving it to our account. If the OCR operation is successful, all the extracted text is stored as a new document otherwise Google Docs will store our original image without any modification. With Google Docs, we can perform OCR on images and PDFs as large as 2 MB, in the output format of Google docs are ODT, PDF, TXT, RTF, DOC and HTML. It supports 30 languages [11], the output text result [11] is represented in Figure 2.8.

```

Algorithm 1 The algorithm of direct k-means clustering
Input: N objects in the R-dimensional space. and M
(number of sub-cluster means)
Output: Initial M sub-clusters that contain all N objects

initialize M first means  $r_1, r_2, \dots, r_M$ 
repeat
for each input object  $x_i, 1 \leq i \leq N$ 
do
assign  $x_i$  to  $D_j$  with nearest mean  $r_j$ , such as
 $\|x_i - r_j\| < \|x_i - r_k\|, \forall k \neq j, 1 \leq j \leq M$ 
for each sub-cluster  $D_j, 1 \leq j \leq M$ 
do
recalculate the mean (centroid) of objects  $x_i \in D_j$ 
If  $r_j = \frac{1}{|D_j|} \sum_{x_i \in D_j} x_i$ , where  $|D_j|$  defines the cardinality
of  $D_j$ 
compute the quality function  $G$ 
until no object has changed sub-clusters (or  $G$  does not
change)
    
```

Figure 2.8 Google Docs

3. COMPARATIVE ANALYSIS

In order to perform the comparative analysis of the OCR tools, this paper consider two performance measures and they are conversion accuracy and error rate. Conversion accuracy is nothing but to identify whether all the alphabets, numbers and special symbols are converted

accurately or not. Error rate helps to identify how much of alphabets, numbers and special symbols are not converted properly. The following tables 3.1 and 3.2 shows the Error rate of OCR tools.

Table 3.1 Comparative Analysis of Online OCR, Free Online OCR, OCR Convert, Convert image to text.net

S. No	Original Text	Online OCR	Free Online OCR	OCR Convert	Convert Image to text.net
1	M	AI	✓	✓	✓
2	first	✓	✓	✓	✓
3	Means	Mmns	✓	✓	✓
4	$r_0^{(1)}, r_0^{(2)}, \dots, r_0^{(M)}$	41), 40, ..., 4m)	r31' . r32). r3")	r\$,", r\$,2).....r\$;"	41), 42), ...or
5	$r_i, 1 \leq i \leq N$	rbiNiNN	ri. IsisN	r,-. I €isN	rb I i-<.Al
6	r_i to D_j	r, to D,	r, to D,	r,- to D,»	ritoDj
7	$r_0^{(1)}$	4P	r8)	1'87	rg)
8	$\varphi(r_i, r_0^{(i)}) \leq \varphi(r_i, r_0^{(i)}), 1 \leq j, u \leq M$	o(c,4ofr,4'1),lNJ, u M	(Mn-18)) S (p(r,-.r8')) .l S j. u S M	<p(r,» .rX)) <<p(r,-.r\$ {'1}).l € j. 11\$ M	49(rhe),54)(rat u), 1j, u<Al
9	D_j	Di	Dj	D,-	Dj
10	$1 \leq j \leq M$	1NjNM	ISjSM	I <j<M	1.<.j4M
11	$r_i \in D_j, r_0^{(i)} = \sum_{i \in D_j} r_i / D_j $	= lea rilipil	rieDPif)" = ZED, r;/D,1	r,-eD',-. rg) = 2,-CD' r,- / D,	rie=EiED, where IDJI
12	$ D_j $	ID,	D,-	D,-	IDJI
13	change	✓	chan e	✓	✓

Table 3.2 Comparative Analysis of Free OCR, i2OCR, Google Docs.

S.No	Original Text	Free OCR	i2OCR	Google docs
1	M	✓	✓	✓
2	first	ï · rst	✓	✓
3	Means	✓	✓	✓
4	$r_0^{(1)}, r_0^{(2)}, \dots, r_0^{(M)}$	r8", r32â€™, ...rï-â€™	r3", r32'.r\$, ""	r3", r32'.r\$, ""
5	$r_i, 1 \leq i \leq N$	r,~, l<:'<N	r,-. I <i<N	r,-. I <i<N

6	r_i to D_j	r , to D ,	r , - to D	r , - to D ,
7	$r_0^{(1)}$	H87	r87	r87
8	$\varphi(r_i, r_0^{(j)}) \leq \varphi(r_i, r_0^{(j)}), 1 \leq j, u \leq M$	$(p(riv' \{P\} \hat{\in} p(riv' \{u\})) \vee lgv u < M$	$\langle p \langle r, .rx^* \rangle \rangle \langle p \langle r, .r; \rangle \rangle .I < 1$. as M	$\langle p \langle r, .rx^* \rangle \rangle \langle p \langle r, .r; \rangle \rangle .I < 1$. as M
9	D_j	D , -	D , -	D , -
10	$1 \leq j \leq M$	$1 < j < M$	$I < j < M$	$I < j < M$
11	$r_i \in D_j, r_0^{(j)} = \sum_{i \in D_j} r_i / D_j $	$r, eD1, rg) = Z, -GD \hat{\in} \text{TM}$ $r, -/ID, - $	$r, -eD \rangle, -.$ $If, \rangle = Z, \rangle, \rangle, \rangle, r, \sim / D_j $	$r, -eD \rangle, -.$ $If, \rangle = Z, \rangle, \rangle, \rangle, r, \sim / D_j $
12	$ D_j $	$D $	D , -	D ,
13	change	✓	✓	✓

Free OCR to Word Convert has obtained the least place because this tool does not produce the accurate results for conversion of characters, symbols and equations. Hence, it is not included in the table. The sample output is given in Figure 2.7.

4. PERFORMANCE MEASURES

The main function of the OCR tools is to convert the given input images into text documents. In order to compare the performance of the above mentioned OCR tools, the following strategies are applied.

Strategy 1: To find the character accuracy (CA) and character error rate (CER) from the resultant text documents. i.e To verify whether an OCR tool has converted all the characters available in the input image correctly or not. For this the following formula is used.

$$\text{Character accuracy (CA)} = (a/n) * 100$$

$$\text{Character error rate} = 100 - \text{CA}$$

(where a=Total number of characters in the resultant text document

n=Total number of characters in the input image)

Strategy 2: To find the special symbols accuracy (SA) and special symbols error rate (SER) from the resultant text documents. i.e To verify whether an OCR tool has converted all the special symbols ($\Sigma, \varphi, \psi, \geq, \leq, =, +, *, -, /, \wedge, \%, \# |, ^{(n)}, (n)$, etc..) available in the input image correctly or not. For this the following formula is used.

$$\text{Special Symbol accuracy (SA)} = (b/m) * 100$$

$$\text{Special Symbol Error Rate SER} = 100 - \text{SA}$$

(where b=Total number of special symbols in the resultant text document

m=Total number of special symbols in the input image)

Table 3 shows the Accuracy and Error rate of different OCR tools. First two columns displays character accuracy (CA) and character error rate (CER), third and fourth column provides special symbols accuracy (SA) and special symbols error rate (SER).

Table 3 Comparison between OCR tools

S.No	OCR Tools	Character Accuracy (CA) (%)	Character Error Rate (CER) (%)	Special Symbols Accuracy (EA) (%)	Special Symbol Error Rate (SER)(%)
1	Online OCR	95.9	4.10	0	100
2	Free Online OCR	98.64	1.36	0	100
3	OCR Convert	100	0	0	100
4	Convert image to text.net	100	0	0	100
5	Free OCR	100	0	0	100
6	i2OCR	100	0	0	100
7	Free OCR to Word Convert	23.29	76.71	0	100
8	Google Docs	100	0	0	100

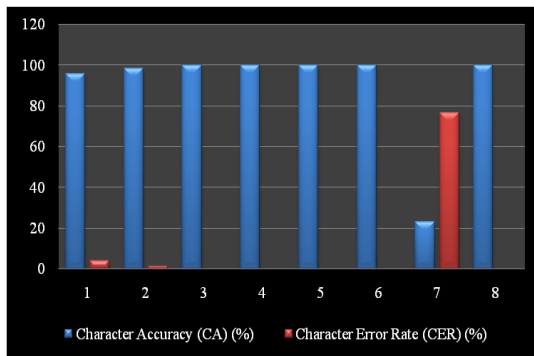


Figure 4.1 Character accuracy and Error rate of the OCR tools

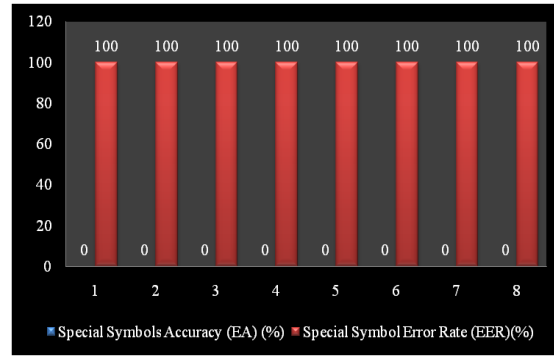


Figure 4.2 Special Symbols accuracy and Error rate of the OCR tools

Figure 4.1 describes the overall character accuracy and error rate of the OCR tools. From this, it is observed the performance of OCR convert, Convert image to text.net, i2OCR and Google Docs are better than other tools. Figure 4.2 presents the overall Special symbols accuracy and error rate of the OCR tools. In this measure all the OCR tools have produced 0% accuracy and error rate for all these tools is 100% which shows no one tool has performed the process of converting the mathematical symbols in equations accurately.

5. CONCLUSION

This work has analyzed the performance of eight different types of OCR tools. From this analysis, we come to know that, the above mentioned OCR software tools cannot detect fonts and formats properly; it gives only plain text as output. Hence, it has proved that the existing OCR tools produced good results for converting characters from the text images but converting

mathematical equations and symbols, these tools have produced inappropriate results. In Future, new OCR tools are to be developed which will be helpful to perform the conversion process appropriately for mathematical equations and symbols.

REFERENCES

- [1] ShivaniDhiman, A.J Singh, "TesseractVsGocr A Comparative Study" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-2, Issue-4
- [2] Chirag Patel, Atul Patel, Dharmendra Patel, "Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study" International Journal of Computer Applications (0975 – 8887) Volume 55– No.10
- [3] http://en.wikipedia.org/wiki/Optical_character_recognition
- [4] <http://www.onlineocr.net/>
- [5] <http://www.newocr.com/>
- [6] <http://www.ocrconvert.com/>
- [7] <http://www.convertimagetotext.net/>
- [8] <http://www.free-ocr.com/>
- [9] <http://www.i2ocr.com/>
- [10] <http://www.ocrtoword.com/>
- [11] <https://docs.google.com/>
- [12] Yasser Alginahi, "Preprocessing Techniques in Character Recognition"
- [13] Oivind due trier, Anil K.Jain, TorfinnTaxt, "Future extraction methods for character recognition A survey".
- [14] Pritpal Singh, SumitBudhiraja, "Feature Extraction and Classification Techniques in O.C.R. Systems for Handwritten Gurmukhi Script – A Survey" Pritpal Singh, SumitBudhiraja / International Journal of Engineering Research and Applications (IJERA), Vol. 1, Issue 4, pp. 1736-1739.
- [15] Youssef Bassil, Mohammad Alwani, "Ocr Post-Processing Error Correction Algorithm Using Google's Online Spelling Suggestion" Journal of Emerging Trends in Computing and Information Sciences, Vol.3, No. 1
- [16] Archana A. Shinde, D.G.Chougule, "Text Pre-processing and Text Segmentation for OCR"IJCSET, Vol2
- [17]https://www.google.co.in/search?q=k+means+algorithm&biw=1366&bih=615&source=lnms&tbm=isch&sa=X&sqi=2&ved=0CAcQ_AUoAmoVChMI5-DkrIvaxgIVD4-OCh0leQqA#imgrc=TQ19POvnV7mNbM%3A
- [18] Sandeep Dangi, Ashish Oberoi, Nishi Goel "Performance Comparison between Different Feature Extraction Techniques with SVM Using Gurumukhi Script" International journal of Engineering Research and Applications(IJERA) ISSN : 2248-9622, Vol. 4, Issue 7(Version 5), July 2014, pp.123-128

BIOGRAPHY

Dr. S. Vijayarani, MCA, M.Phil, Ph.D is working as Assistant Professor in the School of Computer Science and Engineering, Bharathiar University, Coimbatore. Her fields of research interest are data mining, privacy and security issues in data mining and data streams. She has published papers in the international journals and presented research papers in international and national conferences.



Ms. A. Sakila has completed M.Sc in Computer Science. She is currently pursuing her M.Phil in Computer Science in the School of Computer Science and Engineering, Bharathiar University, Coimbatore. Her fields of interest are Data Mining and Multimedia Mining.

