ICDAR2017 Competition on Recognition of Early Indian Printed Documents – REID2017

C. Clausner[1], A. Antonacopoulos[1], T. Derrick[2] and S. Pletschacher[1]

1: Pattern Recognition and Image Analysis Research Lab
School of Computing, Science and Engineering
University of Salford
Greater Manchester, M5 4WT, United Kingdom
www.primaresearch.org

2: Digital Scholarship
British Library
London, NW1 2DB, United Kingdom
www.bl.uk/subjects/digital-scholarship

Abstract—**This paper presents an objective comparative evaluation of page analysis and recognition methods for historical documents with text mainly in Bengali language and script. It describes the competition (modus operandi, dataset and evaluation methodology) held in the context of ICDAR2017, presenting the results of the evaluation of seven methods – three submitted and four variations of open source state-of-the-art systems. The focus is on optical character recognition (OCR) performance. Different evaluation metrics were used to gain an insight into the algorithms, including new character accuracy metrics to better reflect the difficult circumstances presented by the documents. The results indicate that deep learning approaches are the most promising, but there is still a considerable need to develop robust methods that deal with challenges of historic material of this nature.**

Keywords - performance evaluation; page analysis; optical character recognition; OCR; layout analysis; recognition; datasets;

I. INTRODUCTION

The British Library (BL) is currently undertaking a ground-breaking project, Two Centuries of Indian Print [1], to digitise and make available as open access 2,500 early printed Indian books (1785-1909) written in Bengali. Complementary material, the Quarterly Lists, consist of catalogue records for all books published in India 1867 to 1947, will also be made openly available through the project.

Sharing accurate transcriptions of the books will greatly benefit the scholarly research community in performing large-scale analysis of the material to reveal new insights into book and publishing history in India. Much of the material up until now has only been accessible in physical form by visiting the Library.

Page Analysis (here page segmentation, region classification, and text recognition) is a central step in the recognition workflow. Its performance significantly influences the overall success of a digitisation system, not only in terms of OCR accuracy but also in terms of the usefulness of the extracted information (in different use scenarios).

Recent advances using deep learning technologies promise to advance OCR beyond traditional approaches. It is unclear, however, how well such methods cope with historical material where not much training data is available.

This competition was organised in collaboration with the British Library and is a spin-off from a long-standing series of ICDAR page segmentation competitions (the oldest running ICDAR competition since 2001). The aim has been to provide an objective evaluation of methods, on realistic datasets, enabling the creation of a baseline for understanding the behaviour of different approaches in different circumstances. Other evaluations of page segmentation methods have been constrained by their use of indirect evaluation (e.g. the OCR-based approach of UNLV [2]) and/or the limited scope of the dataset (e.g. the structured documents used in [3]. In addition, a characteristic of most competition reports has been the use of rather basic evaluation metrics. While the latter point is also true to some extent of early editions of this competition series, which used precision/recall type of metrics, the 5th edition of the ICDAR Page Segmentation competition (ICDAR2009) [4] made significant additions and enhancements.

This edition (REID2017) is based on the same principles established and refined by the 2011 to 2015 competitions on historical document layout analysis [5] but its focus is on text recognition performance. The evaluation metrics selected for REID reflect the significant need to identify robust and accurate methods for large-scale digitisation initiatives.

An overview of the competition and its modus operandi is given next. In Section III, the evaluation dataset used and its general context are described. The performance evaluation methodology is described in Section IV, while each participating method is summarised in Section V. Finally, different comparative views of the results of the competition are presented and the paper is concluded in Sections VI and VII.

II. THE COMPETITION

REID2017 had three objectives. The first was a comparative evaluation of the participating methods on a representative dataset (i.e. one that reflects the issues and their distribution across library collections that are likely to be scanned). The second objective was a detailed analysis of the performance of each method from different angles. Finally, the third objective was a placement of the participating methods into context by comparing them to open-source systems currently used in industry and academia.

The competition proceeded as follows. The authors of candidate methods registered their interest in the competition and downloaded the *example* dataset (document images and associated ground truth). The *Aletheia* [7] ground-truthing system (which can also be used as a viewer for results) and code for outputting results in the required PAGE format [8] (see below) were also available for download. Two weeks before the competition closing date, registered authors of candidate methods could download the document *images* of the *evaluation* dataset. At the closing date, the organisers received both the executables and the results of the candidate methods on the evaluation dataset, submitted by their authors in the PAGE format. The organisers then verified the submitted results and evaluated them.

Figure 1. Example page images.

III. The Dataset

The importance of the availability of realistic datasets for meaningful performance evaluation has been repeatedly discussed (e.g. [9]) and the British Library selected a subset of current digitisation endeavours. The competition was originally composed of two challenges, but no submissions were made for the Quarterly Lists challenge (recognition of tabular material in both English and Bengali), leaving only the Bengali texts. The corresponding digitisation project at the BL will be digitising 2,500 printed books, amounting to about 500,000 pages in TIFF format. The text of these books is in Bengali language dating between 1785 and 1909. For the most part, the scanned images contain single column lines of text, with a small amount containing illustrations as well as text. Some pages contain marginal data such as numbers, handwritten notes, and decorative frames.

For this competition, the evaluation set consisted of 26 page images as a representative sample ensuring a balanced presence of different issues affecting layout analysis and OCR. Such issues include non-straight text lines, show-through or bleed-through, faded ink, decorations, the presence non-rectangular shaped regions, varying text column widths, varying font sizes, presence of separators and various aging- and scanning-related issues.

In addition to the evaluation set, five representative images were selected as the example set that was provided to the authors with ground truth. Examples from both sets can be seen in Fig. 1.

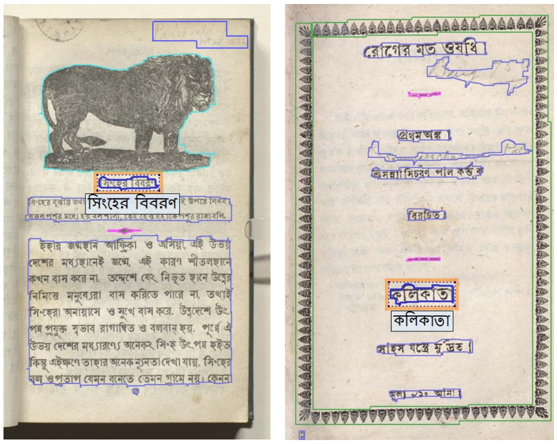

Figure 2. Sample images showing the region outlines (blue: text, magenta: separator, green: graphic, cyan: image) and text content of a selected region.

The ground truth is stored in the XML format which is part of the PAGE (Page Analysis and Ground truth Elements) representation framework [8]. For each region on a page there is a description of its outline in the form of a closely fitting polygon. A range of metadata is recorded for each different type of region. For example, text regions hold information their *logical label* (e.g. heading, paragraph, caption, footer, etc.) among others. Moreover, the format offers sophisticated means for expressing reading order and more complex relations between regions. Sample images with ground truth description can be seen in Fig. 2. The text transcription was provided by the British Library.

IV. Performance Evaluation

A. *Layout Analysis*

The page layout performance analysis method used for this competition [10] can be divided into two main parts. First, correspondences between ground truth and segmentation result regions are determined based on overlapping and missed parts. Secondly, errors are identified, quantified and qualified in the context of different use scenarios.

The region correspondence determination step identifies geometric overlaps between ground truth and segmentation result regions. In terms of Page Segmentation, the following situations can be determined: merge, split, miss / partial miss, and false detection. In terms of Region Classification, considering also the type of a region, an additional situation can be determined: misclassification.

Based on the above, the segmentation and classification errors are *quantified*, recoding the amount of each single error. This data (errors) is then *qualified* by the significance, using two levels. The first is the implicit *context-dependent* significance. It represents the logical and geometric relation between regions. Examples are *allowable* and *non-allowable* mergers. A merger of two vertically adjacent paragraphs in a given column of text can be regarded as allowable, as the result will not violate the reading order. On the contrary, a merger between two paragraphs across two different columns of text is regarded as non-allowable, because the reading order will be violated. To determine the allowable/non-allowable situations accurately, the reading order, the relative position of regions, and the reading direction and orientation are taken into account.

The second level of error significance reflects the additional importance of particular errors according to the use scenario for which the evaluation is intended.

Both levels of error significance are expressed by a set of weights, referred to as an *evaluation profile* [10]. Appropriately, the errors are also weighted by the size of the area affected (excluding background pixels). In this way, a missed region corresponding to a few characters will have less influence on the overall result than a miss of a whole paragraph, for instance.

For comparative evaluation, the weighted errors are combined to calculate overall error and success rates.

B. *Text Recognition*

For the evaluation of OCR results, character-based and word-based measures were used. The former gives a detailed insight into the recognition accuracy of a method while the word-based approach is more realistic in terms of use scenarios such as keyword-based search.

A major problem for the evaluation is the influence of the reading order of text regions. For simple page layouts, the order is obvious, but for more complex layouts, the reading order can be ambiguous. In such cases, measures that are affected by the reading order are less meaningful. An OCR method might recognise all characters perfectly, but if it does not return the regions in the same order as in the ground truth (or with merge/split errors), it will get a very low performance score. Special care was therefore taken when selecting the evaluation measures.

The Character Accuracy [12] is based on the edit distance (insertions, deletions and substitutions) between ground truth and OCR result. The method was extended by the authors to reduce the influence of the reading order. The edit distance is thereby calculated for parts of the texts, starting with good matches and marking matched parts as "visited" until the whole text was processed (unmatched parts count as deletion or insertion errors). The extended measure is called Flex Character Accuracy.

The word-based measure called *Bag of Words* (see [11]) disregards reading order entirely since it only looks at the occurrence of words and their counts, not at the context or location of a word.

Because some of the document pages contain padding characters such as "….." or "- - - -", a preprocessing step is performed to remove special characters from all ground truth and OCR result texts. These include: hyphen, dash, full stop, tilde, asterisk, equal sign, bullet, and double quotes. In addition, unnecessary white spaces are removed (e.g. multiple spaces and trailing line breaks). This helps to focus the evaluation on the more interesting parts of the documents.

All evaluation methods and the datasets are available at the PRImA website [13].

V. Participating Methods

Brief descriptions of the methods submitted to the competition are given next. Each account has been provided by the method's authors and summarised by the organisers.

A. *Google Multi-Lingual OCR*

The Google entry for REID2017 is a small client program that communicates with the publicly accessible Google Cloud Vision API: https://cloud.google.com/vision/. The DOCUMENT_TEXT_DETECTION feature is selected, which instructs the service to expect dense, book-like page images, as opposed to material such as natural scene images. No pre-processing or post-processing is performed by the client program; it relies entirely on the publicly available cloud service for the entire operation. The results submitted for the competition were produced in June 2017. Because the Cloud Vision models get updated periodically, re-running at a later date may produce different results.

Behind the API, the OCR process is split into three phases: text detection, line decoding, and layout analysis.

Text detection locates individual lines of text in the image; these regions are then extracted and provided to the line decoding phase, described below. Text detection follows the approach described by Bissacco et al. [14].

Once detected and extracted, each line sub-image is subject to text recognition by a machine-learned sequence-to-sequence decoder, dubbed "Aksara." For each input line sub-image, the Aksara decoder produces as output a sequence of symbols (including whitespace) along with their bounding boxes. In detail, associated with each image X and hypothesized sequence of Unicode code points Y is a quantity C(X, Y), interpreted as the cost of producing Y given X, and computed as the weighted sum of individual cost components each defined by a single feature function. Two main feature functions are employed: an optical model that operates on pixels, and a character-based language model that encourages linguistic plausibility in the output sequence. Several additional feature functions primarily compensate for sequence-length-related effects related to language model scoring. The weights for combining the cost components are determined via minimum error rate training as proposed by Macherey et al. [15]. Training data comes both from synthesizing textual content from Wikipedia using the Pango text rendering library and from self-labelled data from various sources. Further details of the Aksara decoder can be found in Fujii et al.'s work [16].

Finally, layout analysis is performed to group the individual lines into higher-order structures such as paragraphs and blocks. This phase is mostly heuristic: lines are ordered according to physical position, and two lines are grouped into the same paragraph/block if the gap between them is below a threshold that depends on the detected text size.

B. Bangla OCR I

This layout analysis system was submitted by Tanmoy Nandi & Sumit Kumar Saha, Gnosis Lab, Kolkata, India, Chandranath Adak, School of ICT, Griffith University, Australia, Durjoy Sen Maitra, Decimal Point Analytics, India, and Bidyut B. Chaudhuri, CVPR Unit, Indian Statistical Institute, India. It works with only printed Bangla (or, Bengali) script. Since the REID2017 dataset contains old printed documents, some rigorous preprocessing is required, using following steps:

1. Median filtering on the signal, i.e. considering only the middle values from the sliding window where all the values of the window were sorted numerically.
2. After removing the initial noise, it was found there are few patches which have disconnected parts at the character level. Joining those disconnected regions is mandatory classifying. A modified closing technique with the combination of erosion and dilation was used.
3. Binarisation of the the input signal using Otsu global thresholding technique.

For the text recognition, Google's Tesseract OCR engine [17] was used (with the pre-trained public model of for Bangla). With the help of Tesseract API, the following steps for recognition were performed:

1. Obtaining word segmented classification, line segmented classification and block segmented classification from both the Tesseract LSTM algorithm and old rule-based algorithm.
2. Use of a hierarchical combinational logic to combine the outcomes from all levels of tesseract engine.

In a post-processing step, non-Bengali characters are replaced to Bengali danda, in the output to improve the precision and recall of the overall system. Finally, after the UNICODE conversion, the system generates PAGE XML.

For an input image containing other scripts such as Devanagari or English the method produces erroneous output. Sometimes, removal of very small text components as noise yields erroneous outcome.

C. Bangla OCR II

This method was submitted by the same team as for Bangla OCR I. The OCR system [19] is specially developed for printed Indian scripts like Bengali and Devnagari and it uses a tree structure with each node having a SVM classifier.

Preprocessing is carried out using following steps:

1. Combined local and global adaptive binarisation using a variation of the method of Ntirogiannis et al. [18].
2. The image is segmented using connected component labelling algorithm and morphological operations.
3. Connected component blobs are again checked with threshold values which are statistically calculated against each blob's height, width, area and non-zero component in that area, if they need to merge with another blob. Those blobs are then treated as word image.

The main classification engine of BanglaOCR-II is a feature-based SVM (Support Vector Machine). Here, all the features are handcrafted spatial domain features based on stroke-directions and structural patterns.

A probable "matra" or "headline" location is calculated from each word image using morphological operations. Based on that matra, each word is segmented in three zones (upper, middle, lower), containing multiple complete and broken characters.

The SVM classifier is a two-level tree structure system. The first level is a character/symbol group classifier, which puts high degree of similar shaped characters into one of several groups. At the second level, several classifiers classify individual characters or symbols as a unique class component from that group. Such two-level tree structure approach has several advantages. For example, many middle-zone characters are similar in shape without upper- or lower-zone components or symbols.

After the classification, the classified zonal character portions are merged through a Bengali orthographic and linguistic knowledge-based automated system. Finally, after the UNICODE conversion, the system generates the XML.

D. State-of-the-art Methods

Tesseract OCR 3.04 and 4.0 (alpha version) [17][20] were used for comparison. Because it is known that the open source version of Tesseract uses a very basic internal binarisation, each version was applied both out-of-the-box (with colour image) and with an externally produced binary image (by ABBYY FineReader Engine 11; labelled "B" in the figures). Tesseract 4.0 is based on a long short-term memory (LSTM) approach. No training was required as language models for Bengali and English are available. The PRImA Tesseract-to-PAGE wrapper tool was used to create PAGE XML for Tesseract 3.04. Tesseract 4.0 was executed using the native command line tool and the output was converted from hOCR format to PAGE XML format using the PRImA PageConverter tool.

VI. Results

Evaluation results for the above methods are presented in this section in the form of graphs and, in part, with corresponding tables.

Although the primary focus of this competition is text recognition, the performance analysis of page segmentation and region classification also give useful insights to pinpoint problems and improve the OCR methods. Figure 3. shows the layout evaluation results using a text-focused profile (i.e. errors on non-textual regions are weighted less significantly). Figure 4. shows the breakdown of the different error types of the evaluation measure.

The Bangla OCR II method performs not well because of mainly two reasons: it cannot cope with decorative frames and it produces regions with word granularity. The

Google multilingual OCR performs best, but has the largest proportion of “miss” errors.

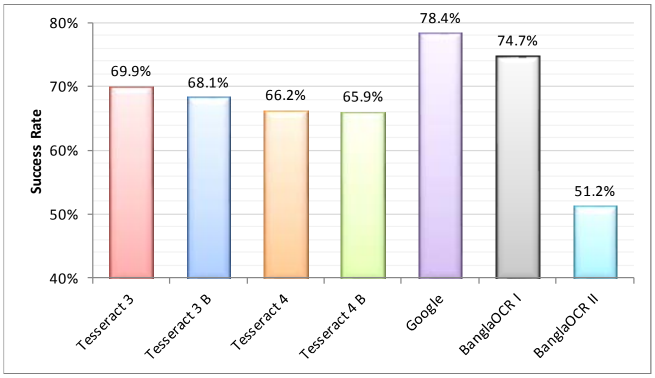

Figure 3. Results using the text region-focused evaluation profile.

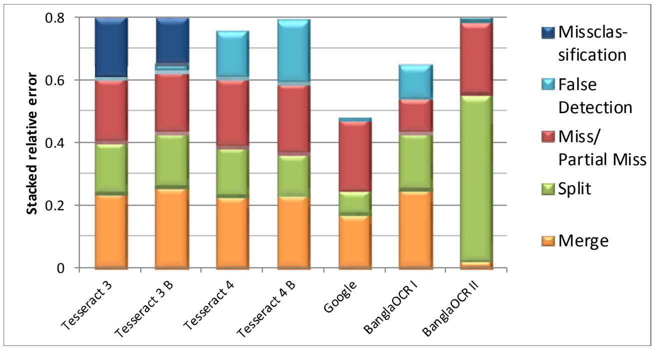

Figure 4. Breakdown of errors made by each method.

Figure 5. shows the traditional and the modified (Flex) character accuracy results. As explained in Section IV.B, there is a clear difference between the two measures, originating from reading order variations. The Flex character accuracy is more meaningful with respect to the actual character recognition. TABLE I. shows the scores for each page.

As mentioned before (Section V.C), the Bangla OCR II method cannot cope with pages containing certain decorative elements. Therefore, a subset of 15 pages without such decorations was evaluated separately (see Figure 6.). It is worth noting that, for the reduced set, Bangla OCR II outperforms both Bangla OCR I and the Tesseract 3 variants.

Considering real-world use cases such as page retrieval via keyword search, a word-based measure is more helpful. Figure 7. shows the results for the Bag of Words measure for all 26 pages and Figure 8. for the reduced set of 15 pages. As can be expected, the success values are lower than the character-based values (one character can cause a whole word to be wrong). The success rate is only based on “miss” errors (words that are in the ground truth but missing or misspelled in the OCR result). False detection (insertion of non-existent words) is disregarded, reflecting the use scenario of page retrieval. The Google multilingual OCR method still outperforms the others, but the margins are narrower.

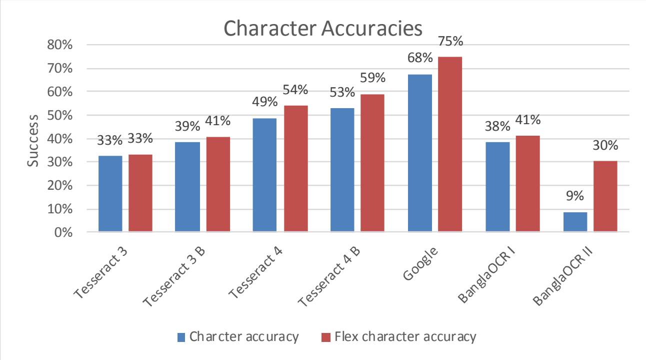

Figure 5. Character accuracy and flex character accuracy (26 pages).

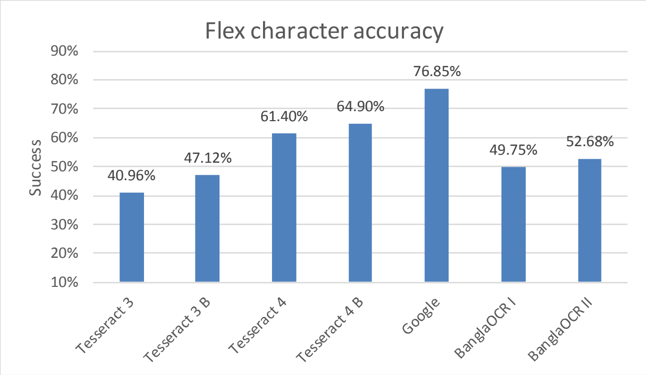

Figure 6. Flex character accuracy for selected pages (15 out of 26).

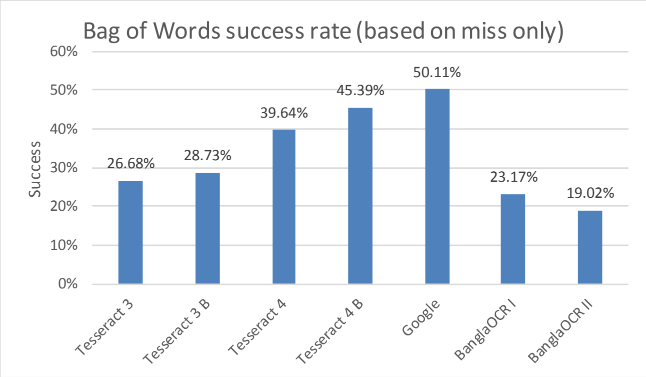

Figure 7. Bag of words success rate (based on miss error, 26 pages).

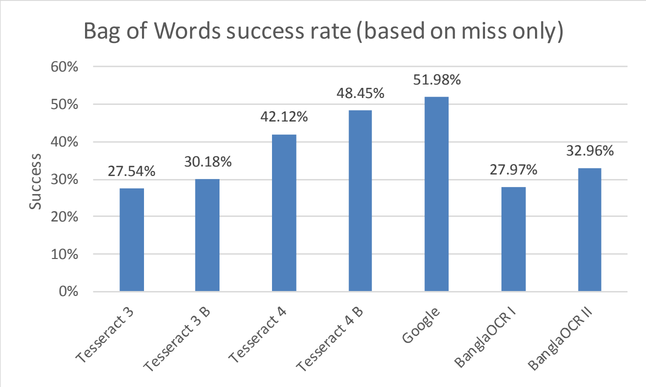

Figure 8. Bag of words success rate (based on miss error, 15 pages).

TABLE I. FLEX CHARACTER ACCURACY PER DOCUMENT (IN %)
(DOCUMENTS WITH DECORATIONS FLAGGED; HIGHEST SCORE IN BOLD)

Decorations	**Tesseract 3**	**Tesseract 3 B**	**Tesseract 4**	**Tesseract 4 B**	**Google**	**Bangla OCR I**	**Bangla OCR II**
Y	28.8	9.0	31.4	42.3	**76.9**	50.0	0.0
Y	18.8	41.2	63.9	58.4	**84.3**	51.8	0.0
N	15.5	31.1	25.4	12.1	44.4	25.4	**62.4**
Y	62.0	64.0	80.5	78.1	**81.1**	61.1	0.0
Y	0.0	0.0	34.0	47.1	**60.9**	0.0	0.0
Y	0.0	2.2	0.0	24.5	**79.6**	0.0	0.0
Y	0.0	33.6	22.4	41.1	**66.4**	0.0	0.0
N	38.3	63.5	79.3	80.9	**85.9**	62.2	68.3
Y	14.2	26.6	31.8	35.1	**71.6**	43.7	0.0
Y	67.2	69.3	75.7	78.9	**86.6**	65.0	0.0
N	44.4	36.4	11.6	15.6	**66.0**	38.8	56.8
N	74.0	70.8	73.7	83.6	**90.4**	59.1	56.2
N	38.7	55.6	48.7	63.6	**67.0**	54.4	52.5
N	33.2	35.4	72.8	77.1	**84.4**	53.2	38.2
N	68.2	70.3	80.2	82.4	**88.5**	63.6	72.1
N	60.2	65.6	74.1	74.6	**79.4**	56.6	69.0
N	49.5	30.9	58.8	63.2	**81.9**	51.5	61.3
Y	50.0	58.5	64.7	69.4	**84.3**	21.4	0.0
N	0.0	0.0	35.6	49.7	**78.0**	0.0	0.0
N	27.7	41.5	70.5	**77.7**	**68.3**	55.9	66.1
Y	0.0	0.0	9.2	9.1	**12.8**	0.0	0.0
N	8.3	27.3	53.2	**61.8**	53.4	42.3	9.3
N	47.0	69.8	83.0	87.3	**88.4**	69.6	65.9
N	44.2	73.6	71.3	79.1	**89.1**	61.5	56.2
Y	11.1	52.1	68.7	79.2	**87.6**	34.2	0.0
N	65.3	35.2	82.8	64.8	**87.7**	52.1	55.9
Avg	**32.6**	**40.4**	**53.2**	**57.8**	**74.5**	**41.2**	**30.2**

VII. CONCLUDING REMARKS

To the best of the authors' knowledge, this competition constitutes the first objective comparative evaluation of page analysis and recognition approaches for historical Bengali documents. It has highlighted the technical difficulties faced by the most advanced methods currently available from academia and industry. The method from Google outperforms the other methods in this instance but there is much room for improvement for all methods. In fact, in certain situations, other methods outperform the Google's method, especially for pages containing a table of content.

A clear first candidate for improvement is the pre-processing stage – especially since the material is of historical nature. This could include a robust binarisation to clearly isolate textual characters and developing a classifier that can handle a variety of historical fonts. A sophisticated approach to recognise both text and decorative elements would also be beneficial. In addition, historical spelling and script variations posed a problem which could be overcome by training and/or dictionary creation in a dedicated project.

REFERENCES

[1] www.bl.uk/projects/two-centuries-of-indian-print, The British Library, accessed 11/07/2017

[2] J. Kanai, S.V. Rice, T.A. Nartker and G. Nagy, "Automated Evaluation of OCR Zoning", *IEEE PAMI,* 17(1), 1995, pp. 86-90.

[3] F. Shafait, D. Keysers and T.M. Breuel, "Performance Evaluation and Benchmarking of Six Page Segmentation Algorithms" *IEEE PAMI,* 30(6), 2008, pp. 941–954.

[4] A. Antonacopoulos, S. Pletschacher, D. Bridson, C. Papadopoulos, "ICDAR2009 Page Segmentation Competition", *Proc. ICDAR2009*, Barcelona, Spain, July 2009, pp. 1370-1374.

[5] C. Papadopoulos, S. Pletschacher, C. Clausner, A. Antonacopoulos, "The IMPACT dataset of Historical Document Images", *Proc. HIP2013*, Washington DC, USA, August 2013, pp. 123-130.

[6] A. Antonacopoulos, C. Clausner, C. Papadopoulos, S. Pletschacher, "ICDAR2013 Competition on Historical Newspaper Layout Analysis – HNLA2013", *Proc. ICDAR2013*, Washington DC, USA, Aug 2013.

[7] C. Clausner, S. Pletschacher and A. Antonacopoulos, "Aletheia - An Advanced Document Layout and Text Ground-Truthing System for Production Environments", *Proc. ICDAR2011*, Beijing, China, 2011.

[8] S. Pletschacher and A. Antonacopoulos, "The PAGE (Page Analysis and Ground-Truth Elements) Format Framework", Proc. ICPR2008, Istanbul, Turkey, August 23-26, 2010, IEEE-CS Press, pp. 257-260.

[9] C. Clausner, C. Papadopoulos, S. Pletschacher, A. Antonacopoulos "The ENP Image and Ground Truth Dataset of Historical Newspapers", *Proc. ICDAR2015*, Nancy, France, Aug. 2015, pp. 931-935.

[10] C. Clausner, S. Pletschacher and A. Antonacopoulos, "Scenario Driven In-Depth Performance Evaluation of Document Layout Analysis Methods", *Proc. ICDAR2011*, Beijing, China, Sept 2011.

[11] A. Antonacopoulos, C. Clausner, C. Papadopoulos, S. Pletschacher, "ICDAR2013 Competition on Historical Book Recognition – HBR2013", *Proc. ICDAR2013*, Washington DC, USA, Aug 2013.

[12] S.V. Rice, "Measuring the Accuracy of Page-Reading Systems", PhD thesis, University of Nevada, Las Vegas December 1996.

[13] PRImA Performance Evaluation Tools http://www.primaresearch.org/tools/PerformanceEvaluation

[14] A. Bissacco, M. Cummins, Y. Netzer, H. Neven "PhotoOCR: Reading Text in Uncontrolled Conditions" in *IEEE Int. Conference on Computer Vision (ICCV)*, Sydney, Australia, Dec. 2013.

[15] W. Macherey, F. J. Och, I. Thayer, J. Uszkoreit "Lattice-based Minimum Error Rate Training for Statistical Machine Translation" in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 725–734, Honolulu, Oct. 2008.

[16] Y. Fujii, D. Genzel, A. C. Popat, R. Teunen "Label Transition and Selection Pruning and Automatic Decoding Parameter Optimization for Time-Synchronous Viterbi Decoding" in Proceedings of the 2015 International Conference on Document Analysis and Recognition (ICDAR), Tunis, Tunisia, November 2015.

[17] R. Smith, "An Overview of the Tesseract OCR Engine," *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, Parana, 2007, pp. 629-633. doi: 10.1109/ICDAR.2007.4376991

[18] K. Ntirogiannis, B. Gatos, I. Pratikakis, "A Combined Approach for the Binarization of Handwritten Document Images", *Pattern Recognition Letters*, vol. 35, pp. 3-15, 2014.

[19] B.B. Chaudhuri, U. Pal, "A complete printed OCR system", *Pattern Recognition*, Volume 31, Issue 5, pp. 531-549, 1998.

[20] Tesseract OCR: https://github.com/tesseract-ocr, accessed 11/07/2017