Legibility and Aesthetic Analysis of Handwriting

Chandranath Adak*, Bidyut B. Chaudhuri[†], Michael Blumenstein*[‡]

*IIIS, School of ICT, Griffith University, Gold Coast-4222, Australia

[†]CVPR Unit, Indian Statistical Institute, Kolkata-700108, India

[‡]School of Software, University of Technology Sydney-2007, Australia

{adak32, bbcisical}@gmail.com, michael.blumenstein@uts.edu.au

Abstract—This paper deals with computer-based cognitive analysis towards legibility and aesthetics of a handwritten document. The legible text creates a human perception that the writing can be read effortlessly because of its orthographic clarity. The aesthetic property relates to the beautiful appearance of a handwritten document. In this study, we deal with these properties on offline Bengali handwriting. We formulate both legibility and aesthetic analysis tasks as machine learning problems supervised by the human cognitive system. We employ automatically derived feature-based recurrent neural networks to investigate writing legibility. For aesthetics evaluation, we employ hand-crafted feature-based support vector machines (SVMs). We have collected contemporary Bengali handwritings, on which the subjective legibility and aesthetic scores are provided by human readers. On this corpus containing legibility and aesthetic ground-truth information, we executed our experiments. The experimental results obtained on various handwritings are encouraging.

Keywords-Aesthetic quality; BLSTM; CNN; Handwriting; Legibility evaluation; SVM.

I. INTRODUCTION

Handwriting is a concatenation of graphical symbols drawn by pen (or, other writing instruments) by the hand in order to represent linguistic constructs for communication and knowledge storage. These graphical marks/writing symbols have deep orthographic relation to the phonology of a spoken language [1]. These writing symbols and orthographic rules evolve through the progress of civilization. The transformation has also occurred with the type of writing tool (from wedge to the pen) and writing medium (from stone-wall to paper).

The purpose of handwriting is the communication between the reader and writer [2]. To achieve this purpose, the writer should know how to write the graphical symbols and the reader should know the phonology about these symbols. Furthermore, one basic criterion, required for proper communication, is the "legibility". According to the Oxford dictionary[1] "legibility" means "the quality of being clear enough to read". The *legibility of handwriting* is a well-known topic in the *Educational research*[2] domain for more than 100 years [3], [4], [5]. But in automated *Document*
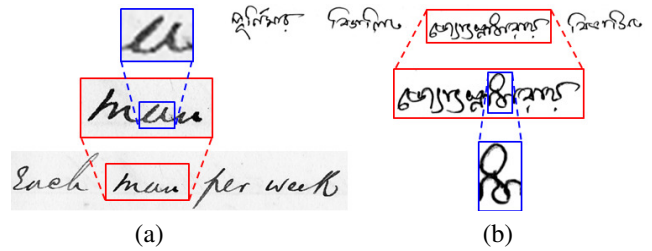
Figure 1.   Examples of (a) English and (b) Bengali readable texts having character-formation ambiguity.

Image Analysis (DIA) and *Handwriting Recognition* (HR) [6] domain, the legibility is relatively a new research issue.

Sometimes legibility is considered synonymous to the "readability" [3]. In Oxford dictionary, the meaning of "readability" is given as "the quality of being legible or decipherable". Readability is actually a higher level of intelligence. It includes legibility with an easy understanding of the semantically well-formed text. For example, if a person, who is well-versed in the English language, but not in Greek, can go through the Greek handwriting written in English alphabet, but cannot comprehend its linguistic meaning. Therefore, for him/her there is void in readability, although the text is legible. We consider the legibility as a part of readability, and if a well-formed text is legible, then it is obvious that the text is readable, provided the reader has linguistic knowledge.

A human reader having usual knowledge of English cursive writing can easily read the text-line in Figure 1(a) as "Each man per week". However, reading the red-boxed highlighted word alone is likely to be confused as 'm'+'u'+'n' or something else, instead of "man", since the blue-boxed character has shape ambiguity. Nevertheless, a human brain, trained with language/script vocabulary and grammar, can make use of neighboring characters/words to get over the ambiguity, and here the human cognition system perceives the word as "man". Similarly, Figure 1(b) shows an example in Bengali script, where the red-boxed readable word contains the blue-boxed ambiguous character. In this paper, although we focus mainly on this alpha-syllabary Indic script Bengali (or, Bangla) [7], [8], we show an example in alphabetic Latin script English in Figure 1(a) for more readerships.

Actually, the legibility depends highly on the character drawing that is based on personal writing style. The ambiguous character formation may lead to poor legibility. Often, this formation of characters is considered as one of the main judgment criteria in the *excellence of handwriting* [4]. The character formation style may be broadly categorized into the *manuscript* and *cursive* [9] writing style. The manuscript writing, where each character is distinctly written, is typically more legible than cursive which is continuous [2].

The manuscript style for the English script is taught at the elementary school level and the cursive style is practiced before the end of primary grades [2]. The Bengali manuscript style is also trained in elementary school level. However, there is no common practice to transit from manuscript to cursive styles in the schools of countries with Bengali mother tongue such as Bangladesh and India. Still, the cursive style is observed in Bengali handwriting, possibly due to the influence of Muslim and British colonial rule [10], when Persian and English scripts dominated in office works. Now a person attains his/her own cursive style by observing the handwriting of teachers, friends, neighbors and some eminent persons, e.g. *R.N. Tagore*.

There are about 300 basic and conjunct characters in Bengali script [8]. Moreover, some conjunct characters can be written in more than one form [11]. Even some educated person may write a complex conjunct character in a wrong way due to lack of proper handwriting education in school. In such scenario, those conjunct characters may individually become illegible. Thus, the research on Bengali handwriting legibility analysis become more challenging owing to the inadequate writing practice in school, the variation of individual cursive writing styles, and the complex patterns of conjunct characters.

Besides legibility, sometimes the handwriting aesthetic receives the attention of the reader. As per the Oxford English dictionary, "aesthetic" is "concerned with beauty or the appreciation of beauty". The aesthetic study of the digital photographic image [12] becomes popular where the researchers attempt to find "beautiful attributes" [13]. In the DIA (*Document Image Analysis*) domain, such aesthetic analysis is also very new [14].

We consider "aesthetic" as a characteristic that gives the perception of beauty to the human cognitive system at the first glance, which may attract a person to read it. Generally, the aesthetic handwriting is more likely to be legible, if honestly written. However, the converse is not generally true. In Figure 2, we present a pair of examples of the handwritten document with high and low aesthetic property.

The purposes of automated analysis of handwriting legibility and aesthetic sense are as follows:
(i) Attempting to impart the machine a human-like perception and to make an artificial cognitive system that understands legibility and aesthetics of handwriting.
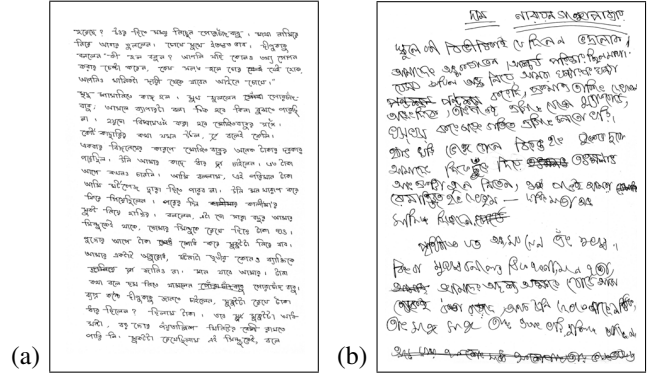*(ii)* Using as a pre-processing module of an OCR (*Optical*

Figure 2. Examples of (a) high and (b) low aesthetic handwritten Bengali document. (The softcopy, after zooming out, exhibits better display).

Character Recognition) engine, for sorting the texts with respect to legible and aesthetic score. It is more likely that highly legible and more beautiful texts will show high recognition accuracy by OCR.
(iii) Analyzing the legibility issues of certain script-specific characters and its influence on aesthetics of the script. For example, people usually have the problem of writing some Bengali conjunct characters properly and beautifully.

Here, we formulate our tasks of legibility and aesthetic analysis as feature-based classification problems. For legibility analysis, the features are extracted automatically using CNN (*Convolutional Neural Network*). To analyze the aesthetics of handwriting, we use some hand-crafted features. The auto-derived and hand-crafted features are fed to RNN (*Recurrent Neural Network*) and SVM (*Support Vector Machine*) classifier, respectively. We employ several human volunteers to put some subjective score as ground-truth.

Our contributions in this paper are as follows:
(i) investigating legibility of handwriting.
(ii) examining the aesthetic value of a handwritten text.
(iii) attempting to find relatively new patch selection strategy without normalizing the input image, for automated feature extraction using CNN.
(iv) analyzing handwriting legibility using a hybrid neural net model (CNN and RNN).
(v) generating a corpus of offline Bengali handwriting with legibility and aesthetic score set by human volunteers.

Very few somewhat related works on this domain exist in the literature of DIA. Chou and Yu [16] considered the quality of offline Chinese character for sorting the characters in a database. They used a feature-based Bayes' decision rule for this sorting. Three features were calculated from character skeleton: (i) stroke-density distribution in four directions, (ii) histogram by horizontal and vertical projection, (iii) stroke crossing-count in horizontal and vertical directions. It was an early attempt for character sorting in terms of character formation quality, somewhat related to legibility.

Majumdar et al. [14] worked with various coarse and fine

level features for visual aesthetic analysis of offline English handwritten document. They employed a linear SVM to classify the documents into five classes, grouped by subjective aesthetic score. They obtained the coarse-level features from word bounding box height, word-spacing, horizontal profiles, and discrete Fourier transformation. Among fine-level features, they used connected component and stroke-based features, gradient and binary descriptor-based features, and texture features. For this fine-level feature extraction, they observed the "local properties of the writing", those were used in "somewhat similar to the degradation models" [14]. However, focusing on handwriting legibility and aesthetics, the assessment of document image quality [15], [17] is a different issue that is related to the degradation due to successive printing (writing), scanning and transmission.

In [18], the fractal behavior of writing strokes was employed to generate some synthetic parameters, with the objective of handwriting classification. With this target, Bouletreau et al. [18] used a parameter named "legibility rating of the image", which is some sort of "fractal dimension of writing". Any direct quantitative analysis of handwriting legibility is missing there due to a different motivation of their work. We did not find any other work fairly related to the automated legibility and aesthetic analysis of handwriting in the domain of DIA. In Sec. III-C, we provide a comparative discussion with these few related works.

The rest of the paper is organized as follows. Sec. II describes the proposed method. Then, the experiments and results are presented in Sec. III. Finally, Sec. IV concludes the paper.

II. Proposed Method

In this section, at first, we describe the problem formulation and then proceed to our planned methodology.

A. Problem Formulation

Both the legibility and aesthetic sense depend on human perception and here we try to teach the machine by this cognitive perceptivity.

Some texts in paragraph, word and character level are shown to the human readers and asked to put a subjective score of legibility within a continuous range of $[L_1, L_2]$; $L_1, L_2 \in \mathbb{R}$. Similarly, for aesthetic text analysis, the human readers are requested to provide some score in a continuous range $[A_1, A_2]$; $A_1, A_2 \in \mathbb{R}$. At the time of data collection, the subjective scores are put in a continuous range, since initially we do not want to restrict the scores in a fixed discrete opinion scale such as *Likert scale* [19].

For each text, multiple persons provide their opinion scores. We calculate the arithmetic mean of all these scores and tag with each text as its *mean opinion score*.

After the data collection, the legibility score range $[L_1, L_2]$ is divided into n_L bins. Therefore, all the *legibility mean opinion score*s of all handwritten texts fall into these n_L bins. In other words, all the handwritten texts are categorized into n_L classes with respect to the legibility score. Similarly, the texts are annotated with *aesthetic mean opinion score*s spreading into n_A bins of score range $[A_1, A_2]$.

Thus, we map the handwriting legibility and aesthetic analysis task into classification problems. Now, the task is to label a given text into n_L-classes of legibility and n_A-classes of the aesthetic score.

B. Preprocessing

In the preprocessing stage, we label the components of a handwritten page by a relatively faster method of single-pass connected component labeling [20]. The text region is separated from the doodle/drawing-like non-text components, if any, by employing the technique in [21]. In the text region, the struck-out texts are also detected using the system of [22]. Very small sized components such as dot, comma, dash, colon etc. and noise are eliminated. The text-lines and words are segmented by employing an off-the-shelf 2D Gaussian filter-based method *GOLESTAN-a*, as discussed in [23]. The character level segmentation is done by the water reservoir principle-based method of [24].

C. Legibility Analysis

For legibility analysis, we require the character level information of a document. Here we employ a feature-based classification method for this analysis.

1) Feature Extraction for Legibility Analysis: For feature extraction, we rely on neural net-based automatically derived features. Here we adopt common LeNet-5 CNN architecture [25] and use it with some modifications as per our requirement, such as the number of feature maps, convolution and sub-sampling operations.

CNN generally takes an input of a fixed size image. For character recognition task, resizing and normalizing the handwritten image work well as input to CNN. However, for a task where writing style holds major information, normalization may impede the performance [26]. In our legibility analysis job, we require the information regarding individual writing style. Therefore, we use the patch-based scheme of [27] with some modification along with CNN architecture, as per our necessity.

For the CNN architecture, we have already obtained the character-level information from the pre-processing step. On a segmented character image, we find its center of gravity (p_{CG}). By centering the p_{CG}, we choose a 128×128 neighbor window, and use this window as a character-level patch input to the CNN.

We do not select the patch through the classical sliding-window technique, since the text-lines are not skew-normalized. Sliding a window through the horizontal median of the text-line main-body height could be performed, but there is high possibility of information loss for extremely skewed/wavy text-lines and less inter-text-line gaps. The
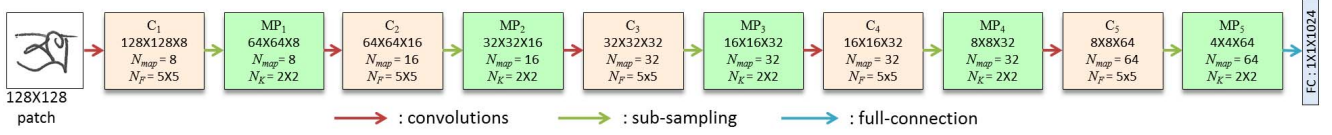
Figure 3. CNN architecture as a feature extractor.

patch-window size is chosen through experimental analysis. It may be noted that a good number of Bengali compound characters are more elongated in vertical than the horizontal direction, but we choose the character-level patch window as a square box. The reason is that we do not normalize the characters, and the character may not fit properly inside a specific sized box. So, sometimes neighboring character components may enter into this box due to improper character segmentation and sometimes a portion of the character may remain out of this box. Thus, we come up with experimentally choosing a 128×128 square window.

This CNN model contains 5 convolutional layers, each followed by a sub-sampling (with *max-pooling* operation) layer. The filter size (N_F), kernel size (N_K) and the number of feature maps (N_{map}) for each of the convolutional layers (C_j, for $j = 1, 2, \ldots, 5$) and subsampling/max-pooling layers (MP_j, for $j = 1, 2, \ldots, 5$) for our CNN architecture are mentioned in Figure 3.

Our automated feature-extractor CNN model ends with a fully connected layer (FC) containing 1024 neurons and thus produces 1024 dimensional feature-vector F_{CNN}.

2) Classification for Legibility Analysis: The extracted features are to be fed into a classifier to assign a class on an unknown handwriting specimen, with respect to legibility score.

We employ a bi-directional *Recurrent Neural Network* (RNN) [28] for the automatically derived CNN-extracted features. The RNN input layer contains an exactly same number of nodes as the dimension of the feature vector, i.e. 1024. The total number of legibility-score classes denotes the number of nodes of output layer. One ϵ node at the output layer is kept extra as null. Two distinct hidden layers for forward and backward sequences are engaged. Here, LSTM (*Long Short-Term Memory*) [29] blocks are employed as hidden units. These two hidden layers comprise 512 and 256 LSTM memory cells, respectively. Such recurrent net prevents the occurrence of so-called "vanishing gradient problem" [30]. The combination of bi-directional RNN and LSTM is briefly called as BLSTM (*bidirectional LSTM*) architecture [30]. All the meta-parameters are tuned and optimized using a tuning set, discussed in Sec. III-B1.

D. Aesthetic Analysis

The aesthetic analysis is performed on the full document as a whole. We do not focus here on character/grapheme-level aesthetic study. Such aesthetic perception is like the visual impact (attraction/distraction) at the first sight. The aesthetic analysis is tackled using a feature-based classification method.

1) Feature Extraction for Aesthetic Analysis: We design some empirical hand-crafted features for the aesthetic analysis of a document.

a) Density of doodle or drawing: A handwritten document may contain non-textual doodle or drawing by a pen/pencil. These doodles/drawings may influence the overall aesthetic perception of the document. Such non-texts are labeled in the pre-processing stage (Sec. II-B). We calculate the black density (d) of a doodle/drawing as follows:

$$d = \frac{number\ of\ black\ pixels}{total\ number\ of\ pixels}.$$

Multiple doodles may be present in a handwritten page. We compute the black densities of all doodles. We also calculate the average density (d_{avg}) and standard deviation of the density (d_{SD}) and use it as feature f_1 and f_2, respectively. If there is no doodle/drawing present in the document, then we set both d_{avg} and d_{SD} as zeros. For the presence of only one doodle in the document image, the value of d_{SD} is zero.

b) Strike-out rate: The presence of struck-out text is possible in a free-form handwritten document. We note that such struck-out text has an impact on the aesthetics of handwriting. These struck-out texts are already detected in the pre-processing stage (Sec. II-B). In a document image, the strike-out rate [22] is calculated and used as feature f_3.

$$Strike\text{--}out\ rate(f_3) = \frac{number\ of\ struck\text{--}out\ words}{total\ number\ of\ words}.$$

c) Marginal space width: The overall aesthetic of a handwritten document depends on the margin space width and its variation. The proper boundary of a text gives a perception of the visual object inside a frame.

The top and bottom marginal white space width from the page borders are calculated using horizontal projection profile [31]. The top and bottom marginal space width from the page borders are used as feature f_4 and f_5, respectively.

From the left border of the page, the marginal white space widths (w_{Li}) of the left-most word bounding box of each text-line are calculated. The average and standard deviation of w_{Li} are calculated, and used as feature f_6 and f_7, respectively. Similarly, the average and standard deviation of right marginal white space widths (w_{Ri}) are computed to employ as feature f_8 and f_9, respectively.

d) Text-line skewness: The text-lines, while mostly written in white blank pages, may be skewed from the standard horizontal axis. Such irregular skewness may influence the handwriting aesthetics.

The text-lines are segmented in the preprocessing stage (Sec. II-B). Through the middle of the text-line, an imaginary straight line is considered. The text-line skewness is measured by an angle between this imaginary line and the horizontal axis. Each text-line may be skewed upward or downward making positive or negative angle from the horizontal axis. We calculate the average and standard deviation of all positive skew angles and use as feature f_{10} and f_{11}, respectively. Similarly, for the downward skewness, the average and standard deviation of all negative skew angles are computed and employed as feature f_{12} and f_{13}, respectively.

Sometimes, there may be waviness (ups and downs) in some text-lines. For such a wavy text-line, we construct the middle imaginary line by drawing a straight line which covers the text-line mostly, and then calculate the above features in a similar way.

e) Word slant: The uniformity in word-slants is considered as a beautiful attribute of handwriting. The words have already been segmented in the pre-processing stage (Sec. II-B). The word slant is calculated by the angle between the vertical part of the character and the standard vertical axis. For the word slant calculation, we use the method of [32] which is based on vertical projection profiles and the *Wigner-Ville distribution* [33]. In a handwritten page, the average and standard deviation of the word-slants are calculated and used as feature f_{14} and f_{15}, respectively.

f) Inter-line and inter-word gap: A writer maintains a white-line gap between two consecutive text-lines, and the white-spaces among successive words. Such inter-line and inter-word gaps enhance the handwriting aesthetics. From the text-line and word segmentation information, these gaps are obtained at the preprocessing step (Sec. II-B). The average and standard deviation of the inter-text-line gaps are used as feature f_{16} and f_{17}, respectively. Similarly, the features f_{18} and f_{19} are obtained from the average and standard deviation of inter-word gaps, respectively.

g) Text main-body height: The regularity in text main body height beautifies the handwritten text. Here we actually calculate the main-body height of individual word. The Bengali words can be partitioned into three zones: upper, middle and lower [8], by two imaginary lines. These two lines can be obtained by horizontal projection profile. The middle zone is considered as text main body region, where the maximum number of black pixels resides and the horizontal histogram reaches global maximum [34]. We compute the height of the main body of all words. The average and standard deviation of these text-heights are used as feature f_{20} and f_{21}, respectively.

In this way, we obtain these 21 single-valued features from a handwritten document image.

2) Classification for Aesthetic Analysis: The extracted features are supplied to an SVM classifier to categorize a document according to its aesthetic score.

To work with the hand-crafted features, we use the SVM with an RBF (*Radial Basis Function*) kernel [35] as a classifier, since it works better than MLP (*Multi-Layer Perceptron*), k-NN (k-*Nearest Neighbors*), MQDF (*Modified Quadratic Discriminant Function*) and SVM-linear in handwriting analysis [22], [36]. The SVM-RBF hyper-parameters (γ and C) are required to be tuned to control the decision boundary and to avoid over-fitting [37]. Such parameters are chosen from a tuning set for the optimal performance of the classifier. This step is known as "model selection". For this model selection, the traditional *grid-searching* technique is used [35]. Here, k-fold *cross-validation* is employed on the training set. The choice of grid-searching range, selection of γ and C, and the choice of k are discussed in Sec. III-B2.

III. EXPERIMENTS AND DISCUSSION

In this section, at first, we discuss the databases employed for the experiments.

A. Employed Database

For experimental analysis, we required a handwritten database having legibility and aesthetic score. We did not find any database with such ground-truth. Therefore, we generated our own corpus with legibility and ground-truth information. For this, we gathered some available offline Bengali handwritings. On this handwriting, we requested volunteers to put some subjective score of legibility and aesthetics as discussed in Sec. II-A.

For both legibility and aesthetic subjective scores, we selected the range [0, 10], i.e. $L_1 = 0$, $A_1 = 0$, $L_2 = 10$, $A_2 = 10$ (refer Sec. II-A). This range was divided into 20 bins, i.e. $n_L = 20$, $n_A = 20$. Therefore, both the legibility and aesthetic analysis tasks can be perceived as 20-class classification problems. We denoted class-1 as lowest score, i.e. poor legibility or poor aesthetics, and class-20 as highest score, i.e. best legibility or best aesthetic opinion.

For a handwritten text, at least 200 human readers provided the subjective legibility and aesthetic scores, and the arithmetic mean opinion score was chosen as the golden standard for each text.

To obtain the legibility score, a human volunteer was shown the cropped handwritten character images arbitrarily, since sequential character showing from a word/text-line may suggest about the writing content due to the semantic knowledge (refer to Figure 1). Here, the volunteers were provided the actual character information to remove the wrong legibility opinion due to erroneous character formation. For example, if a reader is shown only the blue-boxed character of Figure 1(a), and asked to put the subjective score without providing the character information,

then (s)he generally gives high legibility score perceiving as character 'u'. Therefore, if the reader is asked to deliver legibility score after knowing that it is actually 'a', then (s)he can connote the deviation of character formation and be able to provide the proper legibility opinion score for better ground-truthing. For delivering the aesthetic score, the human volunteers were requested to see the full page writing at a distance from 12-18 inches approximately.

For providing the subjective score continuously, the human eyes may be tired after a certain time and the sensitivity of human cognition may deteriorate. To get rid of this, we gave a break to the volunteer after collecting either the legibility scores of 100 characters/ortho-syllables, or aesthetic scores of 50 handwritten pages.

The legibility opinion scores were collected from the native Bengali volunteers who can read and comprehend Bengali writing. The aesthetic scores were collected from volunteers of two separate groups: (i) G_{Ben}: native Bengali persons, (ii) G_{nonBen}: non-Bengali persons who cannot read a Bengali writing.

The employed Bengali handwritings are as follows:
(i) NewISIdb: HwC [11] (say, D_C): This handwritten character database contains a total of 212,300 Bengali characters of 193 distinct types.
(ii) NewISIdb: HwW [10] (say, D_W): This database contains a total of 118 high-frequency Bengali words written by various writers. Here the total number of words is 47,200.
(iii) NewISIdb: HwP_Basic [10] (say, D_{PB}): This database contains 300 copies of handwritings of a Bengali paragraph, written by multiple writers. This paragraph contains all the 50 basic characters (11 vowels and 39 consonants) of Bengali script.
(iv) NewISIdb: HwP_Conjunct [10] (say, D_{PC}): 50 copies of a text document, containing most of the conjunct characters are present in this database. This text is big enough containing 595 words. A single copy usually takes 2-3 pages to write this text. Hence, this database contains total 136 pages of writing.
(v) CMATERdb1.1.1 [38] (say, D_{CM}): This subpart of the database CMATERdb1 [38], contains 100 Bengali handwriting pages.
(vi) ICDAR-2013 Segmentation [23] (say, D_{DAR}): This dataset contains 50 pages of Bengali handwritten document provided during ICDAR-2013 handwriting segmentation contest [23].
(vii) Struck-out Text [22] (say, D_{ST}): This database comprises 250 Bengali handwritten document images containing some struck-out texts. This dataset details can be found in [22].

The datasets D_C, D_W, D_{PB}, D_{PC}, D_{CM} were used for the legibility analysis, and the datasets D_{PB}, D_{PC}, D_{CM}, D_{DAR}, D_{ST} were employed in the aesthetic analysis. The set D_{C1} contained 50% data of D_C and D_{C2} contained remaining 50% data of D_C, i.e. $D_{C1} \cup D_{C2} = D_C$. Similarly, $D_{W1} \cup D_{W2} = D_W$, $D_{PB1} \cup D_{PB2} = D_{PB}$, $D_{PC1} \cup D_{PC2} = D_{PC}$, $D_{CM1} \cup D_{CM2} = D_{CM}$,

Table I
LEGIBILITY ANALYSIS

Set Training \ Test	F-Measure (%)					
	D_{C2}	D_{W2}	D_{PB2}	D_{PC2}	D_{CM2}	D_T' *
D_{C1}	**81.54**	79.65	80.24	68.74	75.73	78.16
D_{W1}	71.63	**78.39**	79.03	61.75	73.91	73.29
D_{PB1}	67.74	72.64	**77.84**	57.33	65.20	67.47
D_{PC1}	78.16	79.27	81.48	**77.12**	76.56	79.47
D_{CM1}	74.36	75.41	79.24	60.85	**78.26**	73.28
$D_{C1} + D_{W1}$	82.34	80.18	80.74	70.48	76.83	79.26
$D_{C1} + D_{W1} + D_{PB1}$	82.56	80.79	81.66	70.93	77.31	78.75
$D_{C1} + D_{W1} + D_{PB1}$ $+ D_{PC1}$	84.79	84.27	86.83	79.65	81.68	84.43
$D_{C1} + D_{W1} + D_{PB1}$ $+ D_{PC1} + D_{CM1}$	85.16	84.90	87.39	79.72	83.17	**85.74**

* $D_T' = D_{C2} + D_{W2} + D_{PB2} + D_{PC2} + D_{CM2}$.

$D_{DAR1} \cup D_{DAR2} = D_{DAR}$, and $D_{ST1} \cup D_{ST2} = D_{ST}$.

B. Results and Evaluation

In this subsection, we present the experimental results and analyze the performance of our system.

1) Legibility Analysis: From the overall engaged database, we used 50% data for training and the rest for testing. More precisely, the sets D_{C1}, D_{W1}, D_{PB1}, D_{PC1}, D_{CM1} were used for training, and the sets D_{C2}, D_{W2}, D_{PB2}, D_{PC2}, D_{CM2} were employed for testing.

The CNN-extracted features were fed to a bi-directional recurrent neural model for legibility analysis. For training purpose, we employed the stochastic gradient descent with a momentum term of 0.9. The initial learning rate was 10^{-3}, decreased by 10% until the stable validation loss. The training epochs were increased from 1,000 up to 40,000.

The legibility analysis performances, in terms of F-Measure, are shown in Table I. We trained the classifier with the different combination of datasets and also tested on various data samples.

Altogether, executing on the total test-set (D_T'), we have obtained 85.74% F-Measure, while training has been performed on entire training-set.

2) Aesthetic Analysis: Here also, we used 50% of the overall employed database for training, and rest 50% for testing. In other words, D_{PB1}, D_{PC1}, D_{CM1}, D_{DAR1}, D_{ST1} were employed as the training set, and the sets D_{PB2}, D_{PC2}, D_{CM2}, D_{DAR2}, D_{ST2} were used for testing.

For aesthetic analysis, the hand-crafted features were fed to the SVM-RBF classifier. The hyper-parameters (γ and C) of the SVM-RBF were tuned using the training set. The best performance was obtained for $\gamma = 2^4$ within the range $[2^{-3}, 2^{-2}, \dots, 2^6]$ and $C = 2^3$ within the range $[2^{-3}, 2^{-2}, \dots, 2^7]$. Here 5-fold cross-validation was used.

The Table II and Table III present the aesthetic analysis performance (in terms of F-Measure) on subjective opinion scores, collected from G_{Ben} and G_{nonBen}, respectively.

Overall, by executing on the entire test-set (D_T''), after training by the whole training-set, we have obtained 86.69%

Table II
AESTHETIC ANALYSIS BY G_{Ben}

Set	F-Measure (%)					
Training \ Test	D_{PB2}	D_{PC2}	D_{CM2}	D_{DAR2}	D_{ST2}	D_T'' *
D_{PB1}	**84.68**	80.92	82.48	79.67	74.48	80.38
D_{PC1}	81.40	**83.79**	83.02	80.66	75.95	81.82
D_{CM1}	82.76	79.45	**84.42**	80.28	76.36	81.26
D_{DAR1}	82.73	80.04	82.39	**83.27**	75.14	81.33
D_{ST1}	75.49	77.18	76.58	76.76	**80.33**	76.78
D_{PB1} + D_{PC1}	85.47	84.68	83.86	81.28	76.17	83.25
D_{PB1} + D_{PC1} + D_{CM1}	86.93	84.74	84.28	81.97	76.38	82.89
D_{PB1} + D_{PC1} + D_{CM1} + D_{DAR1}	87.45	84.89	85.62	84.07	76.55	84.60
D_{PB1} + D_{PC1} + D_{CM1} + D_{DAR1} + D_{ST1}	87.82	86.18	85.91	84.58	81.78	**86.69**

* D_T'' = D_{PB2} + D_{PC2} + D_{CM2} + D_{DAR2} + D_{ST2} .

Table III
AESTHETIC ANALYSIS BY G_{nonBen}

Set	F-Measure (%)					
Training \ Test	D_{PB2}	D_{PC2}	D_{CM2}	D_{DAR2}	D_{ST2}	D_T'' *
D_{PB1}	**85.63**	78.55	82.37	78.28	73.30	80.28
D_{PC1}	81.36	**81.92**	81.65	79.34	72.17	79.87
D_{CM1}	83.64	80.26	**85.74**	80.38	74.78	81.61
D_{DAR1}	81.74	79.84	80.93	**82.06**	73.66	79.45
D_{ST1}	74.29	75.96	77.69	75.69	**78.64**	77.44
D_{PB1} + D_{PC1}	85.78	82.73	83.28	80.15	73.56	81.85
D_{PB1} + D_{PC1} + D_{CM1}	86.31	83.34	87.13	80.79	74.85	82.78
D_{PB1} + D_{PC1} + D_{CM1} + D_{DAR1}	86.58	84.58	87.27	82.50	74.92	84.18
D_{PB1} + D_{PC1} + D_{CM1} + D_{DAR1} + D_{ST1}	87.24	85.52	87.83	83.72	80.25	**85.33**

* D_T'' = D_{PB2} + D_{PC2} + D_{CM2} + D_{DAR2} + D_{ST2} .

Figure 4. Aesthetic analysis with subjective scores of G_{Ben} and G_{nonBen} groups.

Figure 5. The impact of different hand-crafted features for aesthetic analysis, while testing on D_T'' after training by entire training dataset.

and 85.33% F-Measure while employing the subjective aesthetic scores of G_{Ben} and G_{nonBen}, respectively.

From Table II and Table III, it can be observed that the aesthetic analysis performance is almost similar, while the subjective scores are provided by volunteers of G_{Ben} (native Bengali person) and G_{nonBen} (non-Bengali person) groups. In Figure 4, we show this graphically while the testing is done on D_T'' after multiple training sessions.

In Figure 5, we present the effect of different hand-crafted features for aesthetic analysis employed in an SVM-RBF classifier. Here the testing is performed on entire test-set (D_T''), while the training is executed on the whole training-set. The x-axis of the chart presented in Figure 5 denotes the number of cumulative features, i.e., '1' indicates feature f_1, '2' indicates $f_1 + f_2$, '3' indicates $f_1 + f_2 + f_3$, and so on. The performance of the aesthetic analyzer correlates positively with increasing of the number of features.

C. Comparative Discussion

Among the works reported in [14], [16] and [18], only that of Majumdar et al. [14] can be compared to some extent. The different motivations of the works of [16] and [18] with different outcomes, restrict us to provide any possible comparison.
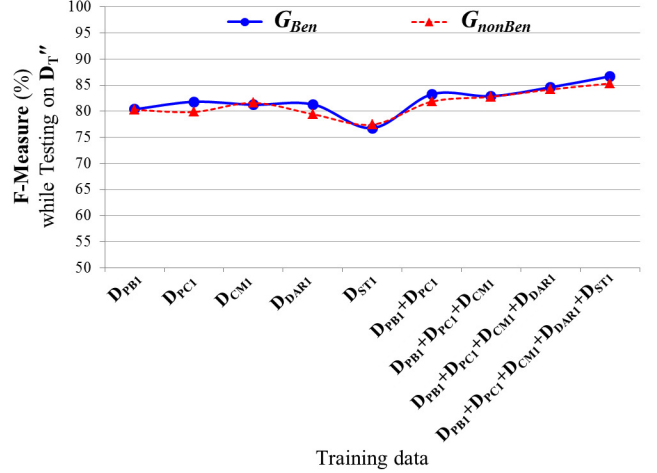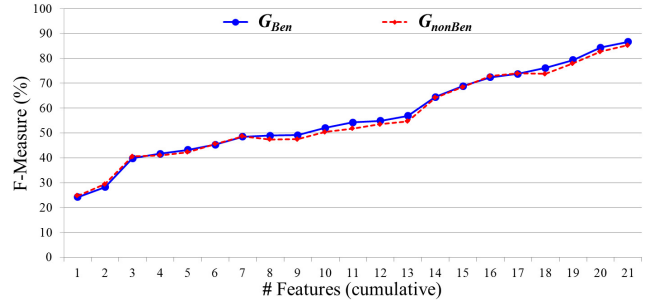
Majumdar et al. [14] worked with an in-house dataset of 1200 English handwritten pages and intended to release 275 document pages with ground-truth (unavailable on their project webpage up to 31st March, 2017). For aesthetics ground-truth generation, they employed a total of only 18 persons and each page was annotated by at least 3 individuals using a discrete 5-point Likert scale. For word-level neatness/fine-level assessment, they obtained 55.88% of best accuracy using SURF (*Speeded-Up Robust Features*). Besides, their system achieved 48.03% best accuracy for page-level neatness/coarse-level assessment employing some statistical features obtained from word segmented image.

We worked with 836 pages of Bengali handwriting for aesthetic analysis. For the ground-truthing, on each page, at least 200 human readers provided the subjective score on a continuous scale. Overall, training on the entire training dataset and testing on the whole test dataset, we obtained 86.69% and 85.33% F-Measure while employing the subjective aesthetic scores given by groups of native Bengali persons (G_{Ben}) and non-Bengali persons (G_{nonBen}), respectively. We also worked on character-level legibility and obtained 85.74% F-Measure while testing on total test-set after training on the entire training-set.

By searching through online and offline archives to the best of our effort, we have not come across any other related work on this problem.

IV. CONCLUSION

In this paper, we work on the legibility and aesthetic analysis from Bengali handwritten document image. These tasks are formulated as feature-based classification problems. We employ the auto-derived CNN extracted features and hand-crafted features for legibility and aesthetic analysis, respectively. The auto-derived features are fed into a recurrent neural network and the hand-crafted features are supplied to an SVM model. By gathering some Bengali offline handwriting, we generate a corpus having ground-truth information of legibility and aesthetic subjective scores. Experimenting on this corpus, we obtain an overall 85.74% F-Measure for legibility analysis. The aesthetic analysis provides an overall 86.69% and 85.33% F-Measure while employing the opinion scores of native Bengali (G_{Ben}) and non-Bengali groups (G_{nonBen}), respectively. In future, we will try to exploit the legibility and aesthetics issues in multi-lingual environments.

REFERENCES

[1] P.T. Daniels, W. Bright, eds., "The Worlds Writing Systems" Oxford University Press, New York, 1996.

[2] E.R. Plattor, E.S. Woestehoff, "The Relationship between Reading Manuscript and Cursive Writing", Elementary English, vol. 44, no. 1, pp. 50-52, 1967.

[3] L. Quant, "Factors Affecting the Legibility of Handwriting", The Journal of Experimental Education, vol. 14, no. 4, Teacher Education, pp. 297-316, 1946.

[4] F.N. Freeman, "An Analytical Scale for Judging Handwriting", The Elementary School J., vol. 15: 8, pp. 432-441, 1915.

[5] D. Starch, "The Measurements of Handwriting", The Journal of Educational Psychology, vol. 4, no. 8, pp. 445-464, 1913.

[6] R. Plamondon, S.N. Srihari, "Online and Off-line Handwriting Recognition: A Comprehensive Survey", IEEE Trans. on PAMI, vol. 22, no. 1, pp. 63-84, 2000.

[7] R.D. Banerji, "The Origin of the Bengali Script", University of Calcutta, 1919.

[8] B.B. Chaudhuri, U. Pal, "A Complete Printed Bangla OCR System", Pattern Recognition, vol. 31: 5, pp. 531-549, 1998.

[9] M.K. Carter, "Manuscript Writing or Cursive Writing?" The Journal of Education, vol. 136, no. 1, pp. 2-3, 1953.

[10] C. Adak, B.B. Chaudhuri, M. Blumenstein, "Offline Cursive Bengali Word Recognition using CNNs with a Recurrent Model", Proc. ICFHR, pp. 429-434, 2016.

[11] C. Adak, B.B. Chaudhuri, "Writer Identification from Offline Isolated Bangla Characters and Numerals", Proc. ICDAR, pp. 486-490, 2015.

[12] R. Datta et al., "Studying Aesthetics In Photographic Images using a Computational Approach", ECCV, pp.288-301, 2006.

[13] L Marchesotti, N. Murray, F. Perronnin, "Discovering Beautiful Attributes for Aesthetic Image Analysis", IJCV, vol. 113, no. 3, pp. 246-266, 2015.

[14] A. Majumdar, P. Krishnan, C. V. Jawahar, "Visual Aesthetic Analysis for Handwritten Document Images", Proc. ICFHR, pp. 423-428, 2016.

[15] P. Ye, D. Doermann, "Document Image Quality Assessment: A Brief Survey", Proc. ICDAR, pp. 723-727, 2013.

[16] S.L. Chou, S.S. Yu, "Sorting Qualities of Handwritten Chinese Characters for Setting Up a Research Database", Proc. ICDAR, pp.474-477, 1993.

[17] V. Govindaraju, S.N. Srihari, "Image Quality and Readability", Proc. ICIP, vol.3, pp. 324-327, 1995.

[18] V. Bouletreau, N. Vincent, R. Sabourin, H. Emptoz, "Synthetic Parameters for Handwriting Classification", Proc. ICDAR, vol. 1, pp. 102-106, 1997.

[19] R. Likert, "A Technique for the Measurement of Attitudes", Archives of Psychology, New York, no. 140, pp. 1-55, 1932.

[20] F. Zhao, H.Z. Lu, Z.Y. Zhang, "Real-time Single-pass Connected Components Analysis Algorithm", EURASIP Journal on Image and Video Processing, vol. 21, pp. 1-10, 2013.

[21] C. Adak, B.B. Chaudhuri, "Extraction of Doodles and Drawings from Manuscripts", Proc. Int. Conf. on Pattern Recognition and Machine Intelligence (PReMI), pp. 515-520, 2013.

[22] B.B. Chaudhuri, C. Adak, "An Approach for Detecting and Cleaning of Struck-out Handwritten Text", Pattern Recognition, vol. 61, pp. 282-294, 2017.

[23] N. Stamatopoulos, B. Gatos, G. Louloudis, U. Pal, A. Alaei, "ICDAR 2013 Handwriting Segmentation Contest", Proc. ICDAR, pp. 1402-1406, 2013.

[24] U. Pal, S. Datta, "Segmentation of Bangla Unconstrained Handwritten Text", Proc. ICDAR, pp. 1128-1132, 2003.

[25] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based Learning Applied to Document Recognition", Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998

[26] A. Schlapbach, H. Bunke, "Writer Identification using an HMM-Based Handwriting Recognition System: To Normalize the Input or Not?", Proc. IGS, pp. 138-142, 2005.

[27] V. Christlein, D. Bernecker, A. Maier, E. Angelopoulou, "Offline Writer Identification using Convolutional Neural Network Activation Features", Proc. German Conference on Pattern Recognition, pp. 540-552, 2015.

[28] M. Schuster, K.K. Paliwal, "Bidirectional Recurrent Neural Networks", IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2673-2681, 1997.

[29] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory", Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[30] A. Graves et al., "A Novel Connectionist System for Unconstrained Handwriting Recognition", IEEE Trans. on PAMI, vol. 31, no. 5, pp. 855-868, 2009.

[31] N. Stamatopoulos, B. Gatos, A. Kesidis, "Automatic Borders Detection of Camera Document Images", Proc. CBDAR, pp.71-78, 2007.

[32] E. Kavallieratou, N. Fakotakis, G.K. Kokkinakis, "Slant Estimation Algorithm for OCR Systems", Pattern Recognition, vol. 34, no.12, pp.2515-2522, 2001.

[33] E. Wigner, "On the Quantum Correction for Thermodynamic Equilibrium", Physical Review, vol.40, pp.749-759, 1932.

[34] U. Marti et al., "Using a Statistical Language Model to Improve the Performance of an HMM-based Cursive Handwriting Recognition Systems", IJPRAI: 15(1), pp.65-90, 2001.

[35] V.N. Vapnik, "The Nature of Statistical Learning Theory", Springer-Verlag, New York, 2000.

[36] C.L. Liu, C.Y. Suen, "A New Benchmark on the Recognition of Handwritten Bangla and Farsi Numeral Characters", Pattern Recognition, vol. 42, no. 12, pp. 3287-3295, 2009.

[37] K. Duan, S.S. Keerthi, A.N. Poo, "Evaluation of Simple Performance Measures for Tuning SVM Hyperparameters", Neurocomputing, vol. 51, pp. 41-59, 2003.

[38] R. Sarkar et al., "CMATERdb1: A Database of Unconstrained Handwritten Bangla and Bangla-English Mixed Script Document Image", IJDAR, vol.15, no.1, pp.71-83, 2012.