New Deep Spatio-Structural Features of Handwritten Text Lines for Document Age Classification

[1]Palaiahnakote Shivakumara, [2]Alloy Das, [3]K. S. Raghunandan, [2]Umapada Pal and [4]Michael Blumenstein

[1]Department of Computer System and Technology, University of Malaya, Kuala Lumpur, Malaysia. shiva@um.edu.my, hudempsk@yahoo.com.

[2]Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India. alloyuit@gmail.com, umapada@isical.ac.in.

[3]Department of Studies in Computer Science, University of Mysore, Karnataka, India. raghu0770@gmail.com

[4]University of Technology Sydney (UTS), Sydney, Australia, michael.blumenstein@uts.edu.au

Abstract: Document age estimation using handwritten text line images is useful for several pattern recognition and artificial intelligence applications, such as forged signature verification, writer identification, gender identification, personality traits identification, and fraudulent document identification. This paper presents a novel method for document age classification at the text line level. For segmenting text lines from handwritten document images, the wavelet decomposition is used in a novel way. We explore multiple levels of wavelet decomposition, which introduce blur as the number of levels increases for detecting word components. The detected components are then used for a direction guided-driven growing approach with linearity, and non-linearity criteria for segmenting text lines. For classification of text line images of different ages, inspired by the observation that, as the age of a document increases, the quality of its image degrades, the proposed method extracts the structural, contrast, and spatial features to study degradations at different wavelet decomposition levels. The specific advantages of DenseNet, namely, strong feature propagation, mitigation of the vanishing gradient problem, reuse of features, and the reduction of the number of parameters motivated us to use DenseNet121 along with a Multi-layer Perceptron (MLP) for the classification of text lines of different ages by feeding features and the original image as input. To demonstrate the efficacy of the proposed model, experiments were conducted on our own as well as standard datasets for both text line segmentation and document age classification. The results show that the proposed method outperforms the existing methods for text line segmentation in terms of precision, recall, F-measure, and document age classification in terms of average classification rate.

Keywords: Handwritten documents, Handwritten text line segmentation, Document quality analysis, Document age estimation, Fraudulent document identification.

1. Introduction

Handwritten document image analysis has received immense attention from researchers because of its numerous real-world pattern recognition and artificial intelligence applications such as in forensics: fake document identification, fraudulent document identification, gender identification, signature verification, personality traits identification, fake answer script identification etc. [1-6]. For all the above-mentioned applications, document age estimation provides a vital clue for finding an accurate solution. Therefore, document age estimation and classification based on the properties of a document is a popular research area in the field of pattern recognition and artificial intelligence. However, when it comes to freestyle writing, the unpredictable effects of degradation due to aging makes the document age estimation process complex and challenging. In addition, document age classification using handwritten text line images is much more complex, and is still considered an open challenge. This is because the existing methods utilize the full page of the document for document age estimation especially for historical document image understanding [7]. Therefore, this study focuses on document age classification at the text line level. It can be noted from the past methods [8, 9] of text line segmentation that when the handwritten document has been affected by several adverse factors, such as degradation, poor quality, touching text lines, etc., segmenting text lines from such documents is challenging.

In the past, several methods [2, 3, 4] have been developed for fraudulent, forged document identification based on studying the quality of the images. However, the primary goal of these methods is to find mismatches and irregularities in writing but not document age classification based on the quality and the effect of text degradations. Therefore, these methods may not be effective for document age classification.

Hence, this work aims at addressing the challenge of document age classification at the text line level. It is observed that as the age of the document changes, the quality of the document decreases. It is evident from the method in [10] that the style of writing changes to a deteriorated form as the age of the document increases. In addition to that, the quality of the document degrades. As a result, the handwritten documents written by people of different ages suffer from inconsistency and irregularity in terms of the font, and the font size. This is due to the change in the speed and force applied to the writing, which results in irregularity in the shapes of the characters. This observation motivated us to study the degree of degradations and character structures in the handwritten text line images of different ages.

In this work, in order to find a solution to this open challenge, and based on the above observations, we explore the topics of image processing, pattern recognition and machine learning as discussed below. The wavelet transform is explored for word component detection in handwritten documents of different ages. Direction-driven region growing with a linearity and non-linearity check, is used for text line segmentation. Structural, spatial and contrast features were used for classification, and the Dense121 architecture was employed for the final document age classification stage. In summary, this illustrates that to solve this complex problem, an intelligent method is proposed consisting of various techniques found in the areas of image processing, pattern recognition and machine learning.

In this work, we consider handwritten documents belonging to different time periods, such as, 1, 3, 6, 12, 18, and 24 months old, for document age classification. For collecting the dataset, we used notebooks and answer scripts stacked in a store room, written by different students at different time periods. The degradations depend on the condition of the store room and the time period the document belongs to. Therefore, document age classification is challenging. The reason to choose the afore-mentioned six classes is as follows. It is noted from the literature that there are methods for degraded document and historical document analysis. These methods work well when the document images contain noticeable degradations. In other words, if the documents contain minute changes due to degradations, the methods may not work well. This is a major flaw of the existing degraded and historical document image analysis based methods. This observation led to consider documents of 1, 3, 6, 12, 18 and 24 months old as a six-class classification problem, because documents at these intervals do not experience material effects in terms of degradation, compared to historical documents that are a few decades old.

The key contributions of the proposed work are as follows. (i) To our knowledge, this is the first work which focuses on classifying handwritten documents according to different ages at the text line level. (ii) Exploring the blurring effect of wavelet decomposition levels for segmenting text lines irrespective of instances of touching text lines, through direction-guided derived boundary-growing and iterative linearity, and non-linearity checking. (iii) Extracting structural, contrast, and spatial features to study the effect of degradations, and to classify documents belonging to different time periods. (iv) Adopting a merger of DenseNet and Multi-layer Perceptron models for successful classifications of text line images of different ages is novel compared to the state-of-the-art methods. (v) The dataset created will be released to the public to support research reproducibility.

The paper is organized as follows. A critical analysis of the existing methods with respect to the proposed problem and the proposed method are presented in Section 2. Section 3 describes the steps for text line segmentation and document age classification. Experimental results and analysis are presented in Section 4. Conclusions and future work are discussed in Section 5.

2. Related Work

Since the proposed document age classification approach involves text line segmentation and forged text detection in document images, we review the methods of text line segmentation from handwritten document images and the methods of fraudulent/forged document identification.

2.1. Handwritten Text Line Segmentation

Romero et al. [13] proposed the influence of text line segmentation in handwritten text recognition. The method works based on zoning such as upper, lower, and middle zones, using projection profiles. However, primarily the method works well for simple document images. Ryu et al. [14] proposed a word segmentation method for handwritten documents based on structured learning, which considers the gap between words and text lines for segmentation. The approach is not effective for touching handwritten documents. Mullick et al. [15] proposed an efficient line segmentation method for handwritten Bangla document images, which uses blurred images for finding seed points that represent spaces between text lines. The scope of the method is limited to a particular script. Kesiman et al. [16] proposed a new scheme for segmenting text lines and characters from gray scale images of palm leaf manuscript, which splits documents into vertical zones to study holes and spacing between characters or text lines. The method is not robust to noise and degraded documents. Garz et al. [17] proposed a user-centered method for segmenting complex historical manuscripts based on document graphs, which captures sparse representation for text line segmentation. The approach does not work for multiple touching lines in the document.

Nhat and Lee [18] proposed dense prediction for text line segmentation in handwritten document images, which explores convolutional neural networks and line adjacency graphs. The method is considered to be computationally expensive. Zhu et al. [19] proposed text segmentation using superpixel clustering, which explores stroke pixels and density based spatial clustering. The method may not work well for noisy and degraded documents. Baig et al. [20] proposed automatic segmentation and reconstruction of historical manuscripts in the gradient domain. The method focuses on separating text from the background in a noisy environment. Vo et al. [21] proposed text line segmentation using a fully convolutional network in handwritten document images. The method explores CNN for segmenting text lines by learning character information and constructing adjacency graphs. The performance of the method depends on the number of samples and large quantity of parameters. Choudhury et al. [7] proposed exploiting force alignment of time–reversed data for improving HMM-based handwriting segmentation. The method explores the Hidden Markov Model for segmenting text lines from handwritten documents. However, the scope of the method is to improve text recognition but not fraudulent document identification.

Li et al. [8] proposed a novel method for text line segmentation for historical document images written in the uchen Tibetan script. The method explores the idea of baseline detection for text line segmentation. However, the scope of the method is limited to a particular type of documents. Kundu et al. [9] proposed text line extraction from handwritten document images using GAN. The method explores the usage of a deep learning model to addresscomplex issues like touching, non-uniformly spaced and multi-skewed text lines. However, it is not sure how well it performs for the documents affected by different conditions. Renton et al. [22] proposed a fully convolutional network with dilated convolutions for handwritten text line segmentation. The method explores deep learning models for learning features and text line segmentation. It can be noted that the method may not work well for arbitrarily oriented text in degraded images.

Li et al. [23] proposed a method for text line segmentation from historical document images, The method uses the information of baselines for segmenting handwritten text lines. Since the method was developed for document images of a specific script, the method is not robust for the documents of other scripts. Gader et al. [24] proposed exploration of deep learning models for text line segmentation from handwritten document images. This method is aimed at a specific type of document image, namely, Arabic handwritten document images. These methods were developed for specific documents written in a specific script, and may not be effective for the documents written in other scripts. Zhou et al. [25] used contour curve tracking for text line segmentation from Tibetan historical document images. The method explores a connected component analysis approach for successful text line segmentation. The scope of the method is confined to a specific script. Roy et al. [26] proposed a Morphology based method for handwritten line segmentation using foreground and background information.

In summary, it is noted from the above discussions that most of the methods directly or indirectly use projection profiles and zoning for segmenting text lines from handwritten documents. However, these methods may not perform well for arbitrarily-oriented text lines and severely touching adjacent lines. There are methods that focus on deep learning models to overcome the problems of the above methods. However, those methods have not been tested on documents that suffer from different levels of degradation.

2.2. Forged Document Detection

Similarly, we review the methods for fraudulent or forged document identification, which are related to document age estimation. Kumar et al. [27] proposed detection of fraudulent alterations in ballpoint pen strokes using a support vector machine, which explores ink of pens for fraud document identification. The features extracted based on ink quality are sensitive to degradations and blur. Halder and Garain [28] proposed a color feature-based approach for determining ink age in printed documents, which uses color for studying the age of a document to identify whether it is old or new. The color features used may not be effective for handwritten document images. Beusekom et al. [29] proposed text line examination for document forgery detection, which explores distortion and misalignment for fraud document identification. The approach is not robust to different aged documents because the extracted forged features overlap with the actual text in the images. Barboza et al. [30] proposed a color-based model to determine the age of documents for forensic purposes, which uses ink quality for finding whether a given document is old or new. The ink quality-based features are sensitive to background changes. Bertrand et al. [31] proposed a system based on intrinsic features for fraudulent document detection, which considers noise and distortion created by forgery operations. The approach works well for noise and degradation-free documents. Ahmed and Shafait [32] proposed forgery detection based on intrinsic document content that uses distortion created by forgery operation. The method is not effective for text lines, as it requires the whole document. Barboza et al. [33] later added strokes or text fraud detection in documents written with ballpoint pens, which studies ink characteristics for exam paper fraud identification. The features are not robust to noise and degradations caused by different ages. Luo et al. [34] proposed localized forgery detection in hyperspectral document images, which uses ink properties for fraud document identification. Khan et al. [35] proposed automatic ink mismatch detection for forensic document analysis by exploring ink properties. The above methods are good for simple or plain documents. Raghunandan et al. [36] proposed Fourier coefficients for fraud handwritten document classification through age analysis. This method explores positive and negative Fourier coefficients for classifying old and new handwritten documents. The method is suitable for a two-class problem but not the multi-class problem as in the proposed work.

Mukhtar & Malhotra [37] proposed a method to identify forgery in handwritten document images. The method uses ink information and extracts statistical features to find matches between the words written using the same. Nandnwar et al. [38] explored Chebyshev-Harmonic-Fourier-Moments and deep learning for forged text detection in handwritten document images. The method obtains a reconstructed image using CHFM to remove all redundant information which is then fed as input to a deep learning model along with the input image for classification of forged handwritten text. Recently, Nandanwar et al. [39] proposed a method for detecting altered texts in document images. The approach explores the combination of positive and negative coefficients of DCT for detecting altered text in the images. The approach may not work well for handwritten documents because it was developed for printed document images.

It is clear from the above discussions that most methods use distortions or noise produced by forgery operations and ink quality either written by pens or printed by printers. The methods target specific document types for achieving better results. A few methods explore the quality of the documents for fraudulent document identification, but none of the methods use quality features for document age estimation. Therefore, document age classification at the text line level is still an open issue. Hence, this paper proposes a new method for document age classification at the text line level based on the combination of hand-crafted features and deep learning models.

3. Proposed Methodology

The proposed method consists of two parts, namely, text line segmentation and document age classification. It is evident that when the document age increases, the quality of documents degrades [40], which results in disconnected character components. Motivated by the work proposed in [41] for character segmentation from video text lines, we propose to explore a similar idea for detecting word components in this work. It is observed that as the number of levels increases in the wavelet domain, the blur also increases, which results in the gap between characters being filled. Therefore, the individual words become distinct components. The proposed method introduces a direction-guided region-growing at the component level through a linearity and non-linearity approach for text line segmentation irrespective of touching between the text lines.

For the second part of document age classification, the text lines given by the first part are taken as input. Due to document aging, there is a high chance of several character components getting split into sub-components. It is evident that this splitting occurs due to document aging and the nature of this phenomenon is heavily dependent on the age of the document. This observation motivated us to extract structural features like the straightness and cursiveness of edge components. Based on straightness and cursiveness, the proposed method classifies the components into different groups. To strengthen feature extraction, for components in each group, the proposed method extracts sharpness [42], contrast, structural and spatial features. Inspired by the special features of DenseNet, namely, alleviation of the vanishing gradient problem, strong feature propagation, feature reuse, and reduction in the number of parameters used, we explore DenseNet121 to classify document images of different ages, namely, 1, 3, 6, 12, 18, 24 months old. The extracted features are fed to a Multi-Layer Perceptron (MLP) model and the input images are fed to the DenseNet121 model. Finally, the proposed network combines the features and the information from the input image for the final classification as shown in Fig. 1. The advantage of DenseNet12 is that when the features of the initial layers fail to extract the finer details in the image due to uneven effects on the quality, the properties of feature propagation and reuse can extract the missing details in the next layers.

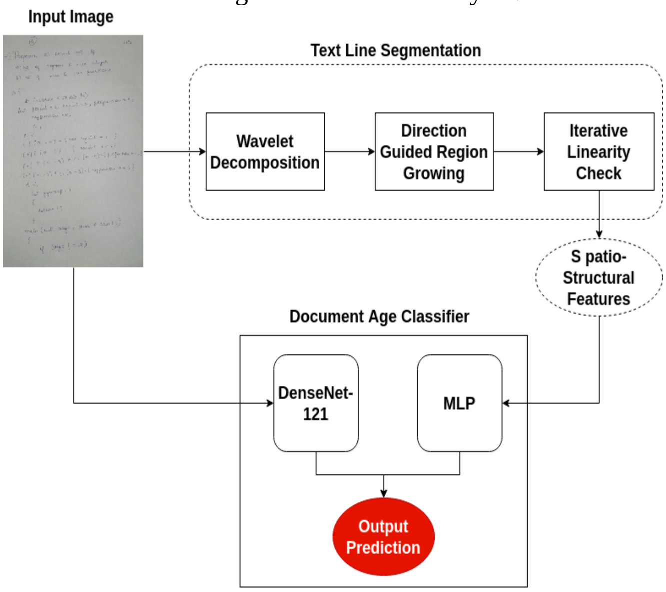


Fig. 1. Block diagram of the proposed work.

Therefore, the proposed method is the combination of handcrafted features and a deep learning model. We believe that the combination is better than the individual approaches because they can be generic, flexible and robust. The problem considered in this work requires such a method because the cause of unpredictable image quality is unknown. In addition, due to free-style handwriting, one can expect large variations in the writing, which makes text line segmentation and document age classification more complex.

For collecting the documents of different ages, we used notebooks and answer scripts written by students in the year 2017. By considering the year 2017 as the reference year, we collected documents which were 1, 3, 6, 12, 18, and 24 months old, i.e., from preceding semesters. For all the documents of the different time periods, the same set of students were asked to rewrite the same content, which was then used to generate a new dataset named as the new class dataset. More details can be found in the experimental section.

3.1. Handwritten Text Line Segmentation

The proposed text line segmentation is divided into three sub-sections, namely Section 3.1.1 which presents a wavelet decomposition-based method for word components detection, Section 3.1.2 describes the direction guided driven boundary growing approach and Section 3.1.3 proposes a linearity and non-linearity check for segmenting text lines.

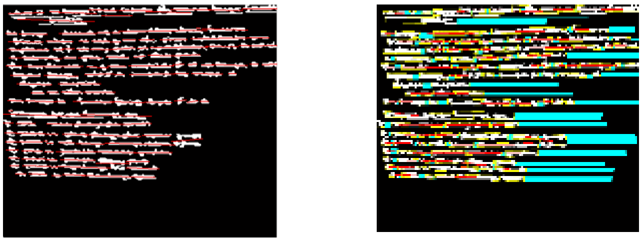
3.1.1. Wavelet Decomposition for Components Detection

For the input image shown in Fig. 2, the proposed method applies wavelet decomposition at the first level, which results in three high frequency sub-bands, namely, horizontal, vertical, and diagonal. It is observed that positive coefficients represent text pixels and negative coefficients represent background pixels. Thus, the proposed method separates positive and negative coefficients from three high frequency sub-band images by displaying white pixels for positive coefficients and black pixels for negative ones as shown in Fig. 2, where H-Level-1, V-Level-1, and D-Level-1 are the results of positive and negative coefficients separation. It is noticed from the results that almost all the text pixels are displayed as white pixels in all the three sub-band images. This shows that the proposed method does not require a binarization method for separating text from background. Since H-Level-1, V-Level-1, and D-Level-1 provide horizontal, vertical and diagonal information, respectively, to obtain the complete information of text, we perform union operation on H-Level-1, V-Level-1 and D-Level-1 as shown in Fig. 2 as Combined-1, where we can clearly see text pixels are grouped as components.

Fig. 2: Wavelet decomposition for components detection

However, we can still observe disconnections due to spaces between character components and word components. Motivated by the work in [41] where it is shown that when we increase the number of wavelet decomposition levels, blur spreads in the same proportion between character components and word components. We increase the wavelet decomposition level until it fills the gap between character components and word components as shown in Fig. 2 as Combined-2, Combined-3 and Combined-4. It can be seen from Combined-2, Combined-3, and Combined-4 in Fig. 2 that as the number of levels increases, the character gaps and word gaps get filled, which results in components. However, finding out the optimal wavelet decomposition level in each case is still an unanswered question. When we look at the combined results in Fig. 2, one can observe that the number of components decreases as the number of levels increases. It is justifiable because as the number of levels increases, character components get merged into word components, and then words components get merged into text line components. As a result, we can conclude that the number of components decreases steeply when word components get grouped as text line components because the whole text line is considered as one single component.

Therefore, using the proposed method, when a sudden change in the number of components detected is observed, it is considered as the stopping criteria. For example, the number of components at the first, second, third, and fourth levels for the input image shown in Fig. 2 are 199, 171, 130 and 29, respectively. There is a big difference between the number of components detected at the third and fourth levels. Therefore, in this work, we consider the components detected at the third level as word components. In other words, the components at the third level are considered as words extracted from the text lines. It is evident from the results shown in Fig. 2 as Combined-3 that all the character components have been grouped as word components.

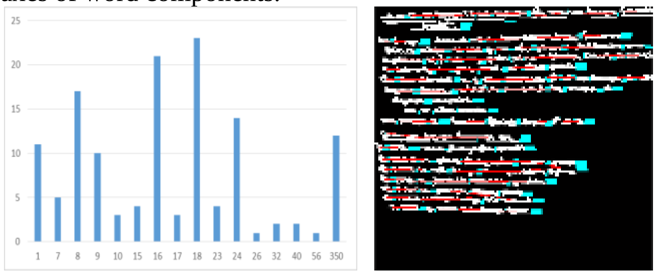


(a) PCA for the words (b) The result of growing

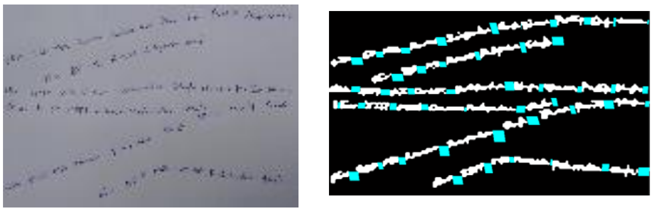
Fig. 3: Growing for grouping the word components using direction given by PCA

3.1.2. Direction Guided Region Growing for Arbitrarily Oriented Text Line Segmentation

The Word component image given by Combined-3 shown in Fig. 2 is taken as the input for text line segmentation. Since the considered document is handwritten, it is not as easy as segmenting text lines from printed documents. As a result, one can expect disconnections in word components, and touching between lines. We believe that in the case of handwritten text, usually all the words in a line are aligned in the same direction with tiny variations in direction. Therefore, we propose to use Principal Component Analysis (PCA) for each word to find the direction of each word. The reason to choose PCA for finding angles of word components is that it is a fact that PCA can give the correct direction if a word misses a few pixels as it considers the majority of the pixels for predicting the direction (principal axis). In the case of handwritten and poor-quality document images, there are chances of losing pixels and shapes. To reduce the effect of these conditions, we prefer PCA for finding the direction of word components as shown in Fig. 3(a), where we can see red marked lines are the principal axes of word components.

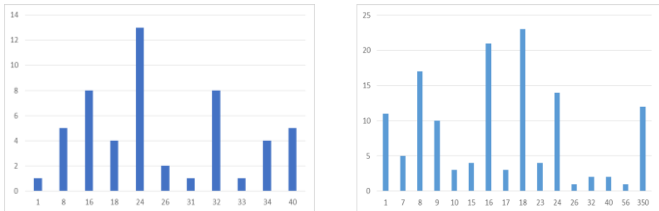


(a) Word gap vs frequencies and the gaps for the text lines

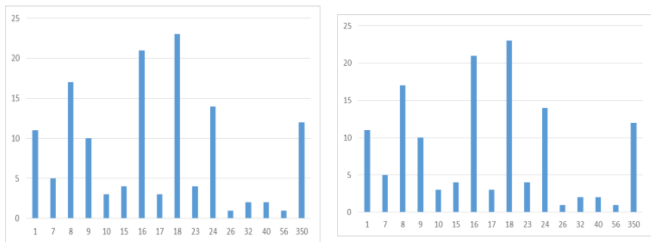


(b) Arbitrarily oriented text lines and respective word gaps (cyan color)

(c) Word gap for 45, 135 and 90 degree rotated documents



Word gap vs frequencies for arbitrary oriented and 45 degree document



(d) Word gap vs frequencies for 135 and 90 degree documents.

Fig. 4: Fixing dynamic threshold for determining end of the line

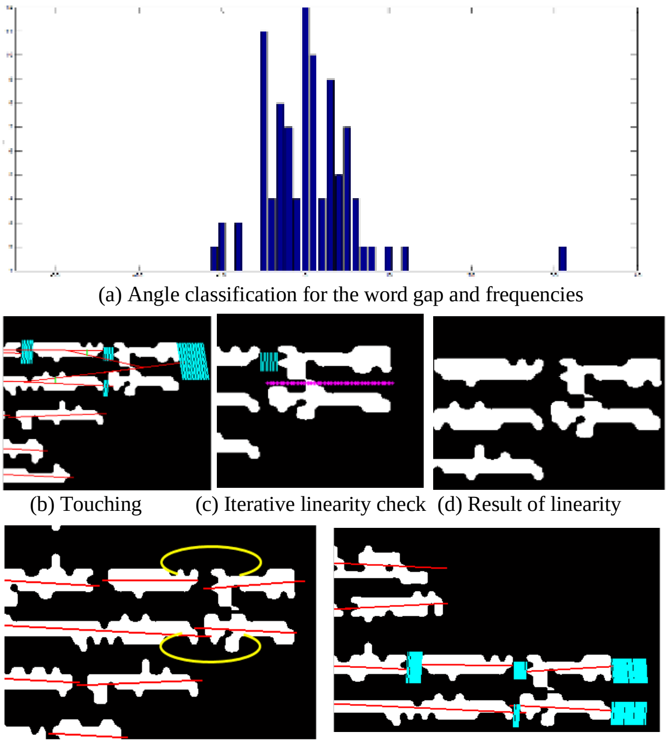
In order to find adjacent word components in the same text line from the current component, in the proposed method we start searching along the word direction until we find the nearest neighbouring word component as shown in Fig. 3(b), where one can see the growing connects all the words in the same line. This results in text line segmentation. The growth is defined as expanding the right or left face of character components of the word component pixel by pixel along the word direction. Fig. 3(b) shows that the growth or expansion stops after reaching the end of the document. This causes a problem when a document contains text lines in different orientations. To overcome this problem, the proposed method finds the gaps between word components in the image irrespective of the orientations of text lines using the growing procedure as explained above. Since the growth connects the nearest neighbouring words, the distance between the current word and the nearest neighbouring word is considered as the word gap.

Then the proposed method performs histogram operation for word gaps as shown in Fig. 4(a), and then chooses the gap, which contributes to the highest peak with a certain constant value as the threshold to end the line. This is valid because the gap between words remains the same for any orientation of text lines. It is illustrated in Fig. 4(b), where it is noted that for arbitrarily oriented text lines in Fig. 4(b), word gaps marked by cyan color are almost the same for all the word gaps. One more illustration for a document of different orientations is shown in Fig. 4(c), where the documents are rotated by 45, 135, and 90 degrees. The histograms of the corresponding rotated document images are shown in Fig. 4(d), from which we can notice that the word gap distance “18” contributes to the highest peak for all the rotations. With these illustrations, we can confirm that the distance between words remains almost the same irrespective of rotations. However, when there is severe touching between the words of two text lines, the direction of the word changes from the text line direction.

3.1.3. Iterative Linearity Checking for Touching Detection

If there exists no touching between the lines, the growth or expansion operation along the word direction works well for text line segmentation. To identify touching, the proposed method checks angles of the current word and the nearest neighbor word given by PCA as discussed in the previous section. If the angle is almost the same (linear), the proposed method allows growing to connect the nearest neighbor word, else it stops at the current word, which results in touching detection. However, the optimal threshold for linearity checking still needs to be determined. As illustrated in the previous section, the distance between words remains the same irrespective of the orientation of the text lines. The same observation led us to estimate the angle between words by calculating the angle of the current word and its nearest neighbor given by PCA.

The proposed method performs histogram operation on angle information as shown in Fig. 5(a), where it can be observed that the “0” degree angle contributes to the highest peak. This is because this angle has been estimated for horizontally oriented document images. This “0” degree is considered as the reference angle. Since this work considers handwritten documents, it is impractical to fix the exact angle of “0” degree, thus we determined the range based on experiments as -15 and +15 from the reference angle as defined in Equation (1). While growing, if the angle between two words lies in this range, it is considered as satisfying the linearity check and the growth continues until the end of the text line. If the angle between words does not lie in the range as shown in Fig. 5(b) due to touching, the proposed approach checks the angle between words iteratively by cutting pixels at the highest deviation angle until it satisfies the defined range as shown in Fig. 5(c). The results of cutting can be seen in Fig. 5(d) and the principal axis of the words after removing touching can also be seen in Fig. 5(d), where the angle between the words satisfies the linearity check. Further, the lines are extracted as shown in Fig. 5(e). This is invariant to rotation because reference angle changes according to the rotation of the document as shown in Fig. 6(a)-(c) for 45, 135, and 90 degrees rotated, respectively. The graphs in Fig. 6 show that for different rotated documents, the iterative linearity check works well.



(a) Angle classification for the word gap and frequencies

(b) Touching (c) Iterative linearity check (d) Result of linearity

(e) PCA axis after cutting (f) Growing based on direction of PCA

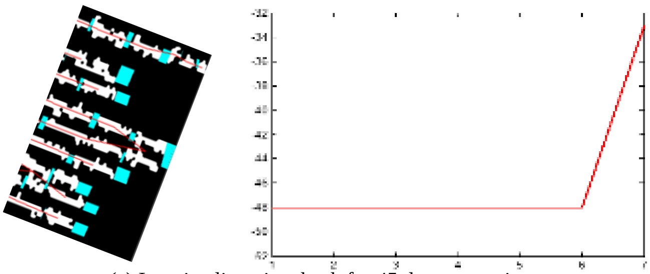
Fig. 5: Automatic threshold for linearity check to solve the touching

The equation for Iterative linearity check, $\mathcal{L}_{check}$, can be given as,

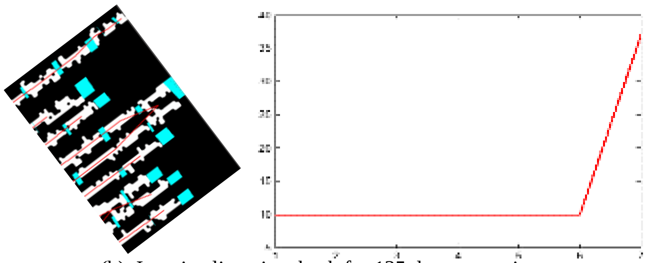
$$\mathcal{L}_{check}(\theta) = \begin{cases} 1, & \theta_d - 15 < \theta < \theta_d + 15 \\ 0, & \theta < \theta_d - 15 \cup \theta > \theta_d + 15 \end{cases} \quad (1)$$

Where θ is the angle for linearity checking, θ_d is the reference angle given by the highest frequency.

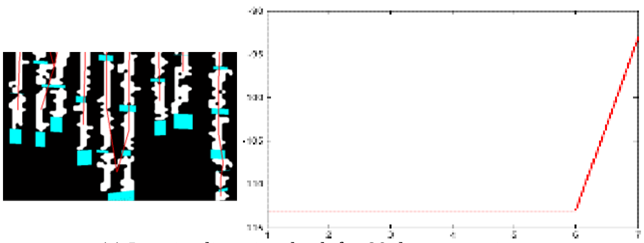
For linearity and non-linearity checks, since the reference angle is calculated automatically based on the spacing between words, it does not affect text line segmentation performance. The values of +15 or -15 are determined empirically after conducting experiments on the chosen random samples. This stimulates tolerance to handle the handwriting variations. The advantage of this step is that it performs well for multiple touching text lines. This is because the direction-guided region-growing and the step of linearity/non-linearity checking work at the component levels. When there is touching between two text lines, the combination of direction-guided region-growing and the linearity-non-linearity check do not satisfy the linearity condition. In this case, the proposed method checks the condition of linearity iteratively by removing the touching region until it satisfies the condition of linearity check.



(a) Iterative linearity check for 45 degree rotation



(b) Iterative linearity check for 135 degree rotation



(c) Iterative linearity check for 90 degree rotation

Fig. 6: Iterative linearity check for different rotations.

3.2. Document Age Classification at the Line Level

The proposed method for document classification is structured into two sub-sections, namely Section 3.2.1 proposes structural features for classifying the components into different categories and Section 3.2.2 presents a method for extracting structural and spatial features from the components in the groups for document classification according to their ages.

3.2.1. Structural Features for Grouping Components

Text lines segmented by the step discussed in the previous section are the inputs for feature extraction. As discussed in the proposed method section, due to degradation, one can expect disconnections, loss of information, and noise introduction during separation of text from background. With this notion, we propose to group edge components of text line images into five sub-groups, which are based on the straightness and cursiveness of edge components. We believe that when a document has good quality, the shapes of edges components are preserved else cursive components turn into straight components due to disconnections. To extract such observation and study the quality of text line images, the proposed method first splits edge components as shown in Fig. 7(a) into sub-components by disconnecting components at the junctions and intersection points as shown in Fig. 7(b), which results in stroke components. For each stroke component in the sub-components image, we propose the following features to group them into five sub-groups.

1) For each stroke component, the proposed method finds the exact middle point by dividing the stroke into two equal parts, and then finds its centroid using the X and Y coordinates of the pixels as shown in Fig. 7(c), where we can see the middle point and centroid are marked for the straight stroke. If the middle point and centroid meet at the same point as shown in Fig. 7(c), the stroke is considered as Exact Straight (ES) as defined in equation (2). The proposed method classifies such stroke components as an ES group.

$$\Delta(a,b) = \frac{l(stroke)}{2}, \qquad C(x,y) = \left[\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right],$$

$$ES = \begin{cases} 1, & \Delta(a,b) \equiv C(x,y) \\ 0, & else \end{cases} \tag{2}$$

where Δ denotes the middle point, C denotes the centroid, l is the length of the stroke, (x_1, y_1) and (a, b) are the coordinates of pixels.

2) The stroke components, which are not classified into ES are considered for further grouping. The proposed approach finds two endpoints for such a stroke component, and then draws lines connecting the two endpoints as shown in Fig. 7(c). If the drawn line intersects at the centroid of the stroke component as shown in Fig. 7(c), the stroke component is considered as Symmetry Curved (SC). The feature is defined in equation (3). The proposed method classifies these components as an SC group.

$$SC = \begin{cases} 1, & \delta\big(\mu(e_1, e_2)\big) \cap C(x,y) \\ 0, & else \end{cases} \tag{3}$$

Where δ is the drawn line using stroke endpoints of $\mu(e_1, e_2)$, e_1 *and* e_2 are endpoints, and $C(x, y)$ is the centroid of stroke μ.

3) For a stroke component which are missed by the above two conditions, if the drawn line between its two endpoints does not intersect with the stroke component (P_{stroke}) as shown in Fig. 7(c), it is considered as Non-Symmetry Curved (NSC) components. The symmetry feature is defined in equation (4). These stroke components are classified as an NSC group.

$$NSC = \begin{cases} 1, & P_{stroke} \neq \delta\big(\mu(e_1, e_2)\big) \\ 0, & else \end{cases} \tag{4}$$

Where δ is the drawn line using the stroke endpoints $\mu(e_1, e_2)$.

4) For the remaining stroke components, as mentioned in the equation defined in (4), if the drawn line does not intersect with the stroke component and creates one hole, it is considered as Curved components as shown in Fig. 7(d). The feature is defined as in equation (5). The proposed approach names the classified components as a C group.

$$C = \begin{cases} 1, & \delta\big(\mu(e_1, e_2)\big) \equiv O \\ 0, & else \end{cases} \tag{5}$$

where δ is the drawn line using the stroke endpoints $\mu(e_1, e_2)$ and O denotes the shape of a hole defined as $O = (\varepsilon_1(\alpha_1, \alpha_2) \in \varepsilon_2(\alpha_1, \alpha_2)$. Here, $\varepsilon_1(\alpha_1, \alpha_2)$ and $\varepsilon_2(\alpha_1, \alpha_2)$ denote the starting and ending points of the stroke, respectively.

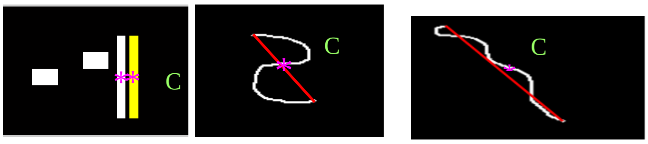
5) For the remaining stroke components, if the centroid falls on the same stroke component as shown in Fig. 7(d), they are considered as a Straight Curved (STC) component as defined in equation (6). Sometimes, the centroid may fall on the same component when it is curved. Therefore, we define the STC components.

$$STC = \begin{cases} 1, & C(x,y) \cap \mu(x,y) \\ 0, & else \end{cases} \quad (6)$$

Where $C(x, y)$ is the centroid of stroke μ and $\mu(x, y)$ denotes the pixel with (x, y) coordinates on the stroke.

(a) Sample Canny edge image of the text line

(b) Split components of the image in (a)



(c) ES SC NSC

(d) Curved STC

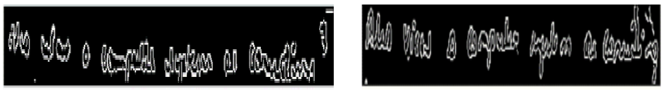
Fig. 7: Straight and cursiveness properties of stroke components

In summary, the proposed method classifies stroke components into five sub-group components according to the above five features. For input text line images of different ages shown in Fig. 8(a) and respective Canny edge images as shown in Fig. 8(b), the effect of grouping can be seen in Fig.8(c) for the ES, SC, NSC, C and STC groups, respectively. It is observed from group-1 to group-5 that stroke components of the groups of new text line images are different from those of the groups of old documents. For instance, when we look at the results of group-1 of the input text line images, groups of the 18-month old document contain more components than the groups of the 1-month old document. It is justifiable because when a document is old, it loses its quality, and hence its components often get split into smaller ones. Therefore, more components satisfy the ES property. The same conclusion can be drawn from the group of NSC, wherein a lower number of components are detected in the group of old text line images as compared to the group of new document images.

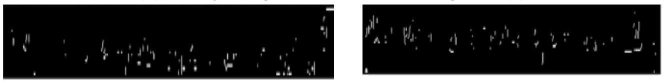
3.2.2. Spatio-Structural Features for Document Age Classification

It is noted that the five groups are obtained by studying the structures of the stroke components as shown in Fig. 8. This section proposes statistical, spatial features, which use stroke components along with contrast, and sharpness features. Therefore, we name them as spatio-structural features as they use both foreground and background information. It is evident from the groups shown in Fig. 8, where we can see the spatial distributions of stroke components of groups belonging to the 1-month and 18 month old text line images, which are different. This observation will be extracted by the spatial features. On the other hand, the sharpness and contrast features extract the quality of the text line images of 1 and 18 month old documents.

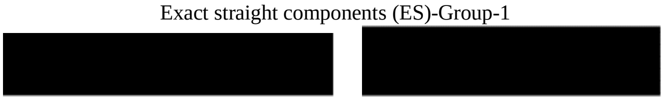
(a) 1-month and 18-months text line images



(b) Canny image of the text line images in (a)

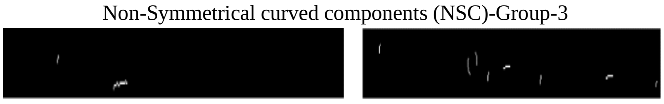


Exact straight components (ES)-Group-1



Symmetrical curved components (SC)-Group-2

Non-Symmetrical curved components (NSC)-Group-3



Curved components (C)-Group-4

Straight-Curved components (STC)-Group-5

(c) Effect of grouping respectively for ES, SC, NSC, C and STC.

Fig. 8: Examples of classifying edge components into groups

The proposed method estimates the proximity distance between pixel to pixel in the groups as shown in Fig. 9(a) and finds the mean/standard deviation for each group. The proposed method then estimates the proximity matrix from the centroid to the centroid of each stroke component in the groups as shown in Fig. 9(b). This results in 5 features. Further, the proposed method finds the centroid and estimates the distance between the centroid to the pixels of the stroke components in the group. This process extracts 5 features. In summary, the proposed method extracts 15 features using spatial information.

Inspired by [42] where Kurtosis has been used for extracting the sharpness of blur images, we propose to use the same Kurtosis method for defining sharpness and to extract the quality of stroke component pixels as defined in equation (10). The proposed method divides the whole group image into eight equal parts. For a pixel of each part, we extract the sharpness, which gives eight sharpness features. Therefore, for 5 groups, a total of 40 features are extracted using sharpness. For the same eight divisions, similar to extraction of sharpness features, we extract the contrast features as defined in equation (11) to study the quality of stroke components. This gives 40 more features. Overall, the proposed method extracts 95 features, which include 15 from spatial information and 80 from quality features.

$$K = \frac{E[(g-\mu)^4]}{E^2[(g-m)^2]} \quad (10)$$

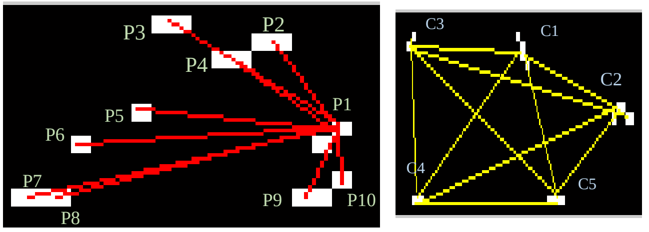
Where the first term is the fourth moment around the mean, divided by the square of the second moment around the mean. The sharpness is estimated as

$$m_k = min\,(ln(K(g_x) + 3)\,, ln(K(g_y) + 3)\,)$$

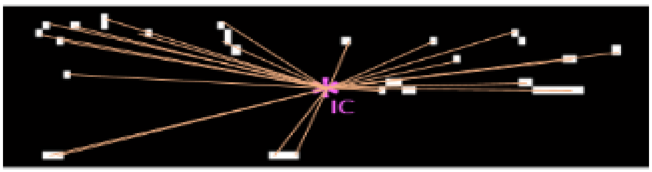
where g_x, and g_y are the gradient magnitudes along the x and y axes, respectively, and

$$MAD(i) = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}|g(x,y) - Median| \qquad (11)$$

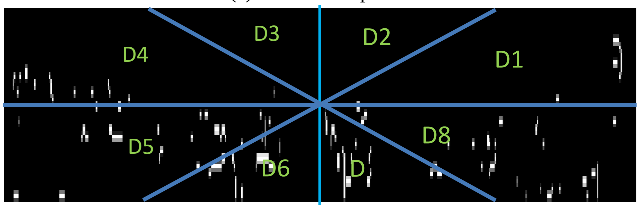
where $g(x,y)$ denotes gradient magnitude along the x and y axes, respectively, and Median is the median of each division.



(a) Pixel to pixels (b) Centroid to centroid

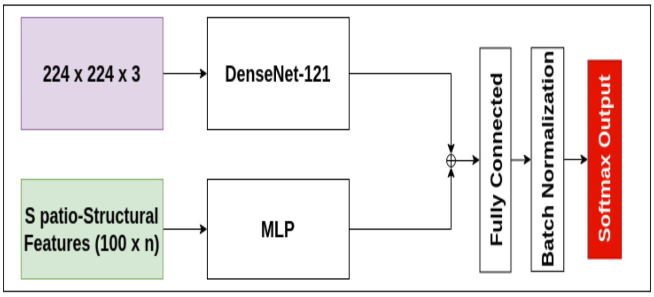


(c) Centroid to pixels

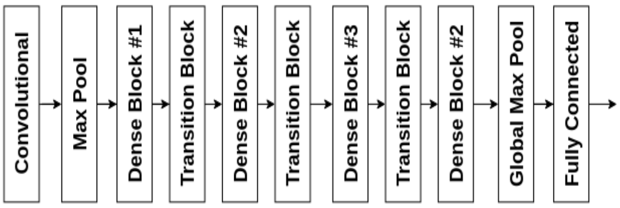


(d) 8 equal divisions for sharpness and contrast features

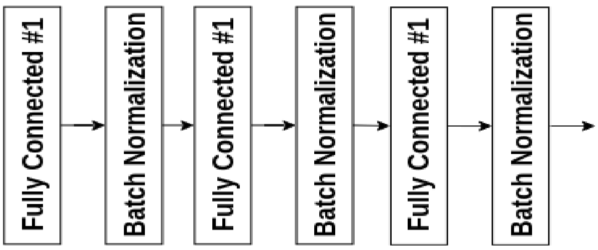
Fig. 9: Spatio-structural feature extraction



(a) Overall Architecture of Document Age Classifier



(b) DenseNet121 Architecture



(c) Multi-layer Perceptron (MLP) Architecture

Fig. 10: Architecture of the proposed network.

3.2.3. Document Age Classification

The extracted 95-dimensional features along with the input image are fed into a two-headed classification network. The architecture of the network is shown in Fig. 10. It accepts two different inputs separately, the document image, and the spatio-structural features of the network. The image, after resizing to 224×224×3, is fed into the classic DenseNet121 network for feature extraction [43]. This network was pre-trained on the Imagenet dataset. The DenseNet architecture comprises a collection of Dense blocks, and Transition blocks apart from other kinds of layers stacked together. A dense block consists of a batch normalization layer, a ReLU activation layer, followed by a convolutional layer repeated multiple times. In the Dense blocks #1, #2, #3, and #4, the set of BN-ReLU-Conv(1×1)-BN-ReLU-Conv(3×3) is repeated 6, 12, 24, and 16 times respectively.

To reduce the size of the feature map as it passes through the network, transition layers are added, which essentially consist of a batch normalization layer, a 1×1 convolutional layer followed by an average pooling layer with a kernel size of 2×2. Finally, after passing through a global max-pooling layer of kernel size 7×7, the feature map obtained is flattened and passed through a fully connected layer, finally resulting in a 1024-dimensional feature set F_1. The spatio-structural features extracted in the previous section are fed into the second head of the network where it passes through a parallel network. This parallel network consists of a multi-layer perceptron with 3 fully-connected layers having 256, 256, and 128 neurons respectively, with a batch normalization layer following each fully connected layer.

The output of this parallel network is a 128-feature set F_2. The two feature sets, F_1 and F_2, are then concatenated and passed on to a fully connected layer followed by the softmax output activation function which gives the probability of the document belonging to each age class considered. The entire network is end-to-end trainable and can be trained once the spatio-structural features of the document have been extracted as described in the previous section.

In this work, the proposed work adopts the predefined Dense121 architecture for document age classification, and we use default parameters, such as ReLU activation functions, Adam optimiser as the optimisation function and a learning rate of 0.0001 for all the experiments.

4. Experimental Results

Since this is the first work that addresses document age classification based on handwritten text lines, there are no standard datasets for experimentation. We created our own dataset for evaluating the proposed method in this work, which includes 6 classes, where the 6 classes include documents which are 1, 3, 6, 12, 18, and 24-month old documents, and has been created using the notes and answer scripts written by university students in the year 2017. It is noted that we used the answer scripts stacked in a storeroom for creating the dataset of different time periods. As a result, it is expected that even a duration of one month might affect the quality of the document images because of dust and other storeroom conditions. Therefore, as the time interval changes or increases, the quality of document images deteriorates. Based on the time periods as mentioned above, the documents are classified as 1, 3, 6, 12, 18, and 24-month old document classes. For the same documents of respective time period classes, we requested the same set of students to rewrite the same content to generate another set of documents which were labeled as a new class.

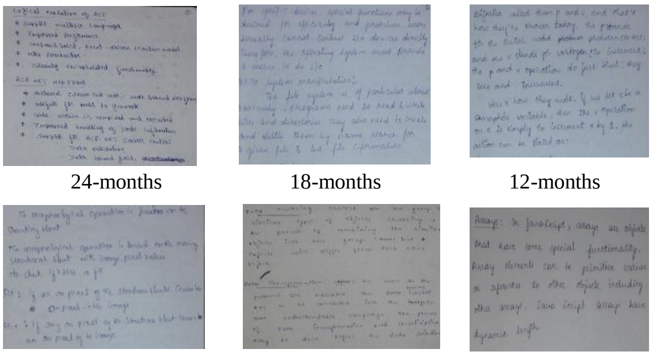
For capturing images, there are no constraints on the use of camera device, gender, script, paper, pen, and the type of ink. As a result, the dataset includes documents written in different scripts, orientations, and written by different genders. Each class contains 100 images, which totals 600 images for experimentation. The reason for choosing the notes and answer scripts from the year 2017 was to preserve privacy and due to the non-availability of data. The same images were used to perform experiments on text line segmentation. To prove the efficacy of the proposed method, for the classification of documents of different ages, we consider a publicly available standard dataset called GOOGLE-LIFE-MAGAZINE (GLM) [28], which consists of documents belonging to five classes of 1930, 1940, 1950, 1960 and 1970 decades. This dataset is also the same as our dataset except the time period difference is in decades.

We use GLM to test the invariant properties of features extracted in this work because these features are extracted at the component level and do not depend on specific text types such as handwritten and printed text. In the same way, the DenseNet121 extracts high level features, namely, edges and context, which are invariant to text type. Therefore, one can infer that the proposed method is robust to different text types.

Furthermore, the images of this dataset contain printed text while the images of our dataset contain handwritten text. Each class contains 40 printed document images, which totals 200 document images. The main objective of considering this dataset for experimentation here is to show that the proposed method is independent of the content in the document. Sample images from our dataset and the GLM dataset are shown in Fig. 11(a) and Fig. 11(b), where we can see handwritten documents of different time periods and GLM images of different decades. It can be seen from the images shown in Fig. 11(b) that the GLM images contain printed text lines embedded on complex background images.

In the same way, to test the robustness of the proposed text line segmentation method, we consider two standard datasets, namely, the ICDAR 2013 robust handwritten text line segmentation contest dataset [44] which includes 150 documents in Greek and English languages, and 50 images in the Bangla language. In total, 200 images are used for experimentation. Out of 200, 50 from each language are considered for testing. Similarly, one more standard dataset, which is publicly available, is considered for evaluating the proposed text line segmentation method [45]. This dataset is much more complex compared to the ICDAR 2013 dataset because it not only contains documents written in different scripts but also in different orientations (left to right, right to left). It is labeled as the PBOK dataset, where P denotes Persian, B denotes Bangla, O denotes Oriya and K denotes Kannada. This dataset consists of 140 Persian, 199 Bangla, 140 Oriya, and 228 Kannada documents. In total, 707 document images are considered for experimentation.

In summary, there are 600 images from our own dataset, 200 from the ICDAR 2013 dataset, and 707 from the PBOK dataset, which totals 1507 different varieties of document images for evaluating the proposed text line segmentation method. For measuring the performance of the methods, we follow the instructions provided in [44, 45] to calculate DR (Recall rate), RA (Precision), and FM (Harmonic mean of Recall and Precision.



24-months 18-months 12-months

6-months 3-months 1-month

(a) Sample images of our dataset

1930 1940 1950 1960 1970

(b) Sample images of Google-LIFE-Magazine dataset

Fig. 11. Sample images of our dataset and the standard Google-LIFE-Magazine dataset.

To compare the result of the proposed text line segmentation method with the results of the existing methods, we implemented the following state-of-the-art methods. Li et al. [23] extracted baseline and edge features for text line segmentation from Uchen Tibetan historical documents. Gader et al. [24] used deep learning models for text line segmentation from handwritten documents. The reason to choose these two methods for a comparative analysis is that their objective of handwritten text line segmentation is the same as the objective of the proposed method. For document age classification, we implemented the following two methods. Mukhtat et al [37] extracted statistical features and ink information for forgery detection in handwritten document images. Nandanwar et al. [38] proposed Chebyshev-harmonic Fourier moments and deep learning for forged text line image classification. The reason for choosing the above forgery detection methods is that the methods analyze the quality of the images for forgery detection, which is similar to the proposed work for document age classification at the text line level.

4.1. Ablation Study for Document Age Classification

The proposed document age classification method at the text line level involves a few key steps, namely, Canny edge, extraction of sharpness features, contrast features, spatial features, and the use of a deep learning model for document age classification. To analyze the contribution of each step, we conducted the following experiments on our dataset and calculated the Average Classification Rate (ACR) for document age classification: (i) For segmented text line images, we used the Canny edge operator to find edge components rather than other edge operators and skeletons. The ACR is calculated using the skeleton of text components rather than the Canny edge of text components. (ii) The ACR is calculated by feeding only the sharpness features to the deep learning model for document age classification. This is to prove the efficacy of the sharpness features for document age classification. (iii) The ACR is calculated by feeding only the contrast features to the deep learning model for classification. This is to test the contribution of the contrast features for document age classification. (iv) The ACR is calculated by feeding only the spatial features to the deep learning model for classification. This is to test the contribution of spatial features for classification; (v) ACR is calculated upon feeding all the features to the deep learning model for classification.

It is noted from Experiments (i)-(iv) that each key step contributes almost equally in the proposed model achieving an ACR of 92.4% as reported in Table 1. Therefore, the key steps are effective. In addition, it is also observed from Table 1 that the individual steps are inadequate to achieve the best results for document age classification, as compared to the proposed method.

Table 1. Analysis of the contribution of the key steps of the proposed method in terms of average classification rate in (%) on our dataset

#	(i)	(ii)	(iii)	(iv)	(v)
Steps	Skeleton	Sharp	Contrast	Spatial	Proposed
ACR	67.9	78.0	64.7	77.8	**92.4**

4.2. Experiments for Text Line Segmentation

Qualitative results of the proposed text line segmentation method on ICDAR, our data, PBOK data, different script data, arbitrarily-oriented text line images, and complex images where touching exists between lines, are shown in Fig. 12. It is observed from Fig. 12 that the proposed text line segmentation method works well for different cases including severely touching text lines.

Quantitative results of the proposed and existing text line segmentation methods for our dataset on each class are reported in Table 2, where one can notice that the value of measures decreases as the age of the document increases. This is understandable because as the age increases, the quality of the document images degrade. It is observed from Table 2 that the proposed method is best at all the three measures for each class compared to the existing methods. However, since the existing methods are limited to specific languages and use language-specific features, the methods report poor results on our dataset compared to the proposed method. On the other hand, the proposed step of text line segmentation does not depend on any language-specific features and is therefore robust to degradations; the proposed method performs the best out of all the methods for text line segmentation.

Table 2. Text line segmentation of the proposed and existing methods on our dataset

Time	Proposed			Li et al. [23]			Gader et al. [24]		
Classes (2017)	DR	RA	FM	DR	RA	FM	DR	RA	FM
1-month	**86.8**	**83.9**	**85.3**	59.9	58.3	59.1	68.1	67.9	68.0
3-months	**85.6**	**82.9**	**84.3**	59.7	58.8	59.3	66.0	64.4	65.2
6-months	**85.4**	**82.0**	**83.6**	58.6	56.3	57.5	64.7	64.3	64.5
12-months	**80.5**	**79.0**	**79.7**	54.3	52.9	53.6	63.9	62.4	63.2
18- months	**79.3**	**77.5**	**78.4**	50.1	49.9	50.0	61.5	60.8	61.1
24-months	**78.1**	**75.4**	**76.8**	50.0	46.2	48.0	59.7	57.4	58.5

Similarly, the results of the proposed and existing methods on the ICDAR 2013 and PBOK datasets are reported in Table 3 and Table 4, respectively. They show that the proposed model is the best compared to the existing methods considered. When we compare the results of the proposed method on the two standard datasets with those of our dataset, it delivers superior results on the standard dataset whereas it delivers lower results on our dataset. The same conclusions can be drawn from the results of the existing methods. This makes sense because our dataset is much more complex compared to the standard datasets. The reason for the poor results of the existing method is the same as discussed above for our dataset. Furthermore, when we look at the results of the proposed method on the ICDAR 2013 and PBOK datasets, the results are consistent for both the datasets while for the existing methods they are not. This shows that the proposed method is reliable and provides stable results for text line segmentation of multi-script and touching text lines.

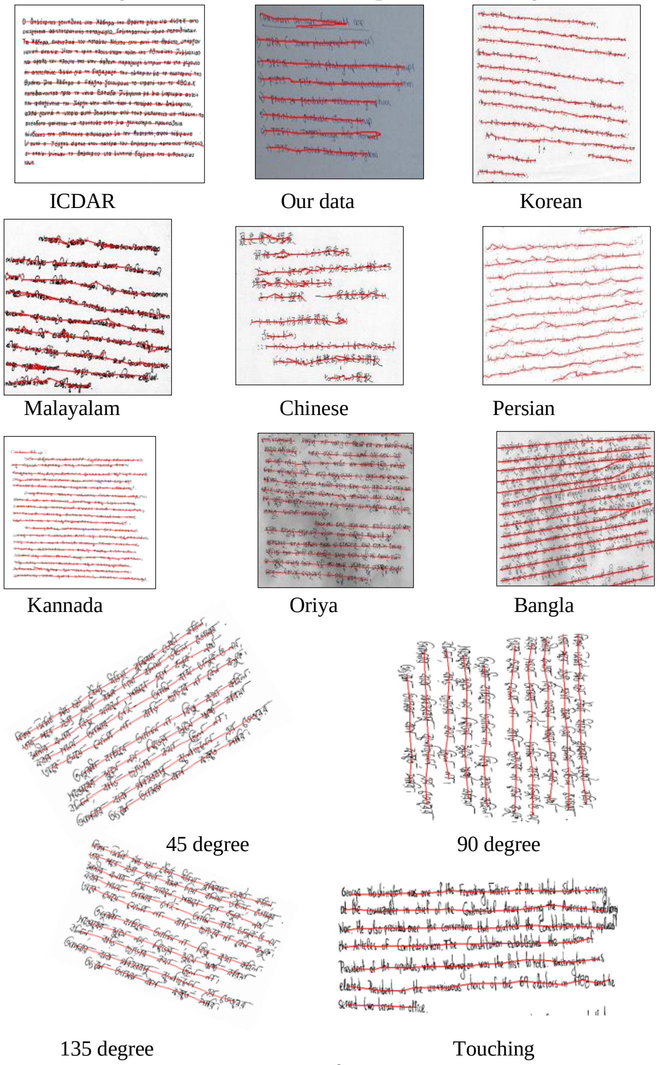


Fig. 12: Qualitative results of the proposed segmentation method for different scenarios.

Table 3. Text line segmentation of the proposed and existing methods on the benchmark ICDAR 2013 robust handwriting contest dataset

Proposed Method			Li et al. [23]			Gader et al. [24]		
DR	RA	FM	DR	RA	FM	DR	RA	FM
97.24	**97.68**	**97.45**	77.12	73.00	75.00	81.00	73.04	75.00

Table 4. Text line segmentation of the proposed and existing methods on the PBOK benchmark dataset

Scripts	Li et al. [23]			Gader et al [24]			Proposed		
	DR	RA	FM	DR	RA	FM	DR	RA	FM

Persian	77.9	75.3	76.6	88.3	86.9	87.6	**94.8**	**94.3**	**94.5**
Bangla	73.2	71.6	72.4	84.4	83.1	83.8	**90.0**	**86.8**	**88.3**
Oriya	77.6	75.4	76.5	84.6	81.9	83.2	**83.8**	**81.5**	**82.6**
Kannada	80.5	78.9	79.7	87.8	85.7	86.7	**96.3**	**95.1**	**95.7**
PBOK	77.3	75.3	76.3	86.3	84.4	85.3	**91.5**	**89.8**	**90.6**

4.3. Experiments for Document Age Classification

Quantitative results of the proposed and existing methods for our dataset and the standard GLM dataset are reported in Table 5-Table 8, where the proposed method outperforms the existing methods for both datasets in terms of the average classification rate. The reason for the poor results of the existing methods is that the successful classification in these methods depends on the distortion created by the forgery operation. In the case of ours and the GLM datasets, the distortion is created due to document aging. These distortions are different from the distortions created by forgery operations, and hence the existing methods deliver poor results. On the other hand, the combination of structural, contrast, spatial features, and DenseNet is capable of dealing with degradations at different levels to achieve the best classification results.

Table 5. Classification rate (in %) of the confusion matrix of the proposed method on our dataset.

2017 Year	1-month	3-month	6-months	12-months	18-months	24-months
1-month	**91.2**	8.5	1.3	0	0	0
3-month	1.8	**89.8**	3.6	2.3	0	0
6-month	1.2	2.7	**88.1**	4.4	3.0	0
12-months	0	0	1.9	**92.5**	4.7	3.1
18-months	0	0	0	3.2	**94.6**	5.4
24-months	0	0	0	0	1.9	**93.1**

Table 6. Classification rate (in %) of the confusion matrix of Nandanwar et al [38] on our dataset.

2017 Year	1-month	3-month	6-months	12-months	18-months	24-months
1-month	**73.9**	4.5	2.1	0	0	0
3-month	12.4	**77.8**	7.8	1.0	3.1	0
6-month	3.3	11.0	**75.3**	8.1	5.8	4.1
12-months	4.1	1.2	10.6	**81.3**	14.2	8.9
18-months	0	1.1	3.7	3.2	**79.9**	9.1
24-months	0	0	0	1.8	3.8	**82.1**

Table 7. Classification rate (in %) of the confusion matrix of Mukhtar et al. [37] on our dataset.

2017 Year	1-month	3-month	6-months	12-months	18-months	24-months
1-month	**64.9**	21.7	23.4	3.3	0	0.0
3-month	9.4	**52.4**	13.2	5.6	3.3	0.0
6-month	20.2	27.3	**55.1**	3.7	19.2	2.2
12-months	16.6	1.5	17.4	**61.8**	25.4	1.1
18-months	3.5	9.6	3.3	18.4	**59.3**	19.2
24-months	0	0	4.1	11.4	10.4	**65.6**

Table 8. Classification rate (in %) of the confusion matrices of the existing methods [38] and [37] on the GLM dataset.

Time gap	Nandanwar et al [37]					Mukhtar et al. [36]					Proposed				
	30's	40's	50's	60's	70's	30's	40's	50's	60's	70's	30's	40's	50's	60's	70's
30's	**69.2**	13.7	7.1	1.7	0.0	**63.5**	16.2	12.1	5.6	0.0	**89.4**	2.1	4.0	0.4	0.0
40's	18.3	**65.3**	4.5	2.4	4.7	19.4	**66.5**	17.7	8.9	1.2	7.1	**85.4**	11.4	0.0	0.0
50's	7.7	18.2	**71.0**	18.4	8.1	15.3	8.9	**68.2**	16.9	15.5	1.0	7.8	**83.9**	2.8	2.36
60's	0.0	3.6	9.9	**69.8**	14.3	1.7	4.9	19.2	**64.8**	11.3	0.0	0.0	1.7	**90.0**	7.2
70's	0.0	0.0	0.0	14.0	**70.4**	0.0	5.2	11.3	12.6	**69.9**	0.0	0.0	0.0	1.3	**88.6**

Although, the proposed method reports the best results, it misclassifies at times due to the following reasons. When the document is severely affected by distortions such as blur and degradations, the text line segmentation may not work well. This is because the severe degradations may result in a larger number of fragments of the text components. When the text component is split into many sub-components, the direction-guided growing may not work well, and hence the linearity and non-linearity check fails to identify touching points. Due to this, the features lose their discriminative power, thus it misclassifies the images.

4.4. Experiments for Evaluating the Robustness of the Proposed Method

As noted from [36], new handwritten document were created for all the documents of different ages and considered as a new class. According to the method [36], the documents of the new classes are considered as "fraud" and the documents of other classes are considered as "original". This is because the fake or the fraudulent documents must have been created after the original one. To test the performance of classification with the new class versus the old classes (1, 3, 6, 12, 18, 24 months old), we considered this problem as a seven class classification problem. For the documents belonging to the old classes, we requested a set of students to re-write the content of the docuents belonging to the old classes at the same time. As a result, the quality of the document images of the new class is better than the quality of the document images of the old classes. The new class contains 600 documents for experimentation. The confusion matrix for the proposed and existing methods for document age classification are reported in Table 9-Table 11, where we can see the proposed method is better than existing methods even when seven classes are considered. The reason for the poor results of the existing methods is that the methods are sensitive to degradations caused by aging. On the other hand, the proposed method is robust against the effects of degradations due to the adoption of structural, contrast, spatial features, and the DenseNet deep learning model.

Interestingly, the classification rate of the new class is higher than the other old classes as reported in Table 9. It is understandable as the quality of the new documents is higher than the documents of the old classes. Overall, the results of document age classification and the seven-class classification demonstrates that the proposed method is independent of the content and the number of classes considered, and hence works well for text line images.

Table 9. Classification rate (in %) of the confusion matrix of the proposed method on our dataset with the new class.

2017 Year	New	1-month	3-months	6-months	12-months	18-months	24-months
New	**90.4**	6.3	1.4	0.7	0.4	0.0	0.0
1-month	8.0	**86.7**	4.5	1.5	0.0	0.0	0.0
3-months	1.9	3.9	**87.3**	3.2	2.2	1.0	0.0
6-months	0.0	0.0	2.7	**88.1**	5.0	3.3	3.1
12-months	0.0	1.5	0.0	2.0	**91.9**	5.1	9.3
18-months	0.0	0.0	2.7	1.9	1.2	**93.1**	5.4
24-months	0.0	0.0	0.0	0.0	0.0	6.8	**92.4**

Table 10. Classification rate (in %) of the confusion matrix of Nandanwar et al. [38] on our dataset with the new class

2017 Year	New	1-month	3-months	6-months	12-months	18-months	24-months
New	**72.8**	5.7	1.4	1.9	0.0	0.0	0.0
1-month	13.4	**73.0**	11.0	7.4	0.8	0.0	0.0
3-months	7.2	12.3	**76.5**	10.2	1.4	0.0	0.0
6-months	0.0	1.2	5.4	**80.4**	7.9	4.9	3.1
12-months	3.4	1.1	2.2	2.3	**81.2**	13.3	9.5
18-months	0.0	0.0	0.6	0.0	3.3	**80.0**	9.0
24-months	0.0	0.0	0.0	1.5	1.4	1.8	**81.0**

Table 11. Classification rate (in %) of the confusion matrix of Mukhtar et al. [37] on our dataset with the new class.

2017 Year	New	1-month	3-months	6-months	12-months	18-months	24-months
New	**63.4**	11.0	17.1	3.9	5.3	0.6	0.0
1-month	11.8	**63.8**	15.6	22.9	2.2	0.0	0.0
3-months	18.9	9.6	**51.9**	15.3	5.9	3.9	0.0
6-months	4.1	18.4	24.7	**55.4**	4.5	20.4	2.5
12-months	5.6	15.4	1.2	15.8	**60.9**	23.8	0.9
18-months	1.6	4.2	9.8	2.4	20.1	**59.6**	21.7
24-months	0.0	0.0	0/0	4.2	10.9	10.0	**66.0**

It is mentioned in the Proposed Methodology Section that the combination of handcrafted features and deep learning models is flexible and robust compared to individual methods because the combined one integrates the merits of feature extraction (generic, flexible and robust) and deep learning approach (accurate and high results when it extracts the relevant features). Therefore, the proposed method is capable of handling a greater number of classes with a few modifications. It is evident from the experimental results reported in Table 9, that these are impressive for seven classes.

However, there are a few limitations of the proposed work. If we decrease the time interval class, the performance of the method degrades because it results in an additional number of classes, and it reduces the quality difference. This leads to poor classification performance. However, this is beyond the scope of the proposed work and it can be considered as our future work. When documents are affected severely by distortions such as blur and degradations, the text line segmentation may not work well because this can have repercussions on the results in more fragments of text components. In this case, the step of text line segmentation fails to segment the correct text lines and therefore the classification step misclassifies the images. In addition, when the handwritten document contains not only handwritten text but also printed text, pictures, equations etc. the proposed method may not work well. With this discussion, one can gather that there is lot of scopes for improving the proposed method in the near future.

5. Conclusions and Future Work

We have proposed a novel method for documents age classification at the text line level. The wavelet decomposition method is used in a novel way for detecting word components from handwritten document images. The components are used for text line segmentation based on a guided-driven boundary-growing approach and iterative linearity, and non-linearity. The proposed method extracts structural, contrast and spatial features from the segmented text lines images. The features, along with the input image, are then fed to a deep learning model which is a combination of DenseNet and an MLP for the classification of images of different ages. The results on ours and the standard datasets show that the proposed method outperforms the existing methods for text line segmentation. Similarly, when we look at the results of classification on ours and the standard datasets, the proposed classification is robust, and generic compared to existing methods. The proposed method can be extended to study personality traits based on quality analysis of text as well as the background of the images.

Acknowledgment

This work received partial support from Faculty Grant (GPF096A-2020, GPF096B-2020 and GPF096C-2020), University of Malaya, Malaysia. Thanks to Tanmay Jain for his sincere help in this work. The authors of this paper also thank B. J. Navya, G. Pooja, University of Mysore, Mysore, Karnataka, India for their support in creating the dataset and implementing the modules.

References

[1] S. K. Jemni, Y. Kessentini and S. Kanoun, “Improving recurrent neural networks for offline Arabic hadnrwitting recognition by combining different language models”, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 34, 2020.

[2] Z. Huang, P. Shivakumara, T. Lu, U. Pal and M. Blumenstein, B. Chetty and G. Hemantha Kumar, “Improved ring radius transform-based reconstruction for video character recognition”, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 35, 2020.

[3] L. Rothacker, F. Wolf and G. A. Fink, “Annotation-free word spotting with bag-of-features HMMs”, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 35, 2020.

[4] S. M. Obaidullh, K. C. Santosh, N. Das, C. Halder and K. Roy, “Handwritten Indic script identification in multi script document images: A survey”, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 32, 2020.

[5] J. Chapran, M. C. Fairhurst, R. M. Guest and C. Ujam, “Task-related population characteristics in handwriting analysis”, IET- Computer Vision, 75-87, 2008.

[6] U. Garain and B. Halder, Even bid data is not enough: need for a novel reference modeling for forensic document authentication”, IJDAR, 1-11, 2019.

[7] Z. Li, W. Wang, Y. Chen and Y. Hao, “A novel method of text line segmentation for historical document image of the uchen Tibetan”, Journal of Visual Communication and Image Representation, 61, 23-32, 2019.

[8] H. Choudhury, S. Mandal and S. R. M. Prasanna, “Exploiting forced alignment of time-reversed data for improving HMM-based handwriting segmentation”, Expert Systems with Applications, 121, 158-169, 2019.

[9] S. Kundu, S. Paul, S. K. Bera, A. Abraham and R. Sarkar, “Text line extraction from handwritten document images using GAN”, Expert Systems with Applications, 140, 2020.

[10] V. Basavaraj, P. Shivakumara, D. S. Guru, U. Pal. T. Lu and M. Blumenstein, “Age classification using disconnectedness features in handwriting”, In Proc. ICDAR, 1131-1136, 2019.

[11] I. Ahmad and G. A. Fink, “Handwritten Arabic text recognition using multi-stage sub-core-shape HMMs”, IJDAR, 329-349, 2019.

[12] K. Adam, A. Baig, A. Bouridane and S. E. Menshawy, “KERTAS: dataset for automatic dating of ancient Arabic manuscripts”, IJDAR, 283-290, 2018.

[13] V. Romero, J. A. Sanchez, V. Bosch, K. Depuydt and J. D. Does, “Influence of text line segmentation in handwritten text recognition”, In Proc. ICDAR, 536-540, 2015.

[14] J. Ryu, H. Koo and N. I. Cho, “Word segmentation method for handwritten documents based on structural learning”, IEEE. SPL, 1161-1165, 2015.

[15] K. Mullick. S. Banerjee and U. Bhattacharya, “An efficient line segmentation approach for handwritten Bangla document image”, In Proc. ICAPR, 2015.

[16] M. W. A. Kesiman, J. C. Buric and J. M. Ogier, “A new scheme for text line and character segmentation from gray scale images of Palm leaf manuscript”, In Proc. ICFHR, 325-330, 2016.

[17] A. Garz, M. Seuret, A. Fischer and R. Ingold, “A user centered segmentation method for complex historical manuscripts based on document graphs”, IEEE Trans. HMS, 181-193, 2017.

[18] Q. Nhat and G. Lee, “Dense prediction for text line segmentation in handwritten document images”, In Proc. ICIP, 3264-3268, 2016.

[19] Y. Zhu and K. Zhang, “Text segmentation using superpixel clustering”, IET Image Processing, 455-464, 2017.

[20] A. Baig, S. A. S. Al-Maadeed, A. Bouridane and M. Cheriet, “Automatic segmentation and reconstruction of historical manuscripts in gradient domain”, IET Image Processing, 502-512, 2018.

[21] Q. N. Vo, S. H. Kim, H. J. Yang and G. S. Lee, “Text line segmentation using a fully convolutional network in handwritten document images”, IET Image Processing, 438-446, 2018.

[22] G. Renton, Y. Soullard, C. Chatelain, S. Adam. C. Kermorvant and T. Paquet, “Fully convolutional network with dilated convolutions for handwritten text line segmentation”, IJDAR, 177-186, 2018.

[23] Z. Li, W. Wang, Y. Chen and Y. Hao, “A novel method of text line segmentation for historical document image of the uchen Tibetan”, Journal of Visual Communication and Image Representation, 61, 23-32, 2019.

[24] T. B. A. Gader and A. K. Echi, "Unconstrained Handwritten Arabic Text-lines Segmentation based on AR2U-Net”, In Proc. ICFHR, pp. 349-354, 2020.

[25] F. Zhou, W. Wang and Q. Lin, “A novel text line segmenation method based on contour curve tracking for Tibetan historical documents”, International Journal of Pattern Recognition, Vo. 32, 2018.

[26] P. P Roy, U. Pal, J. Lladós, "Morphology based handwritten line segmentation using foreground and background information", In Proc. ICFHR, pp. 241-246, 2008.

[27] R. Kumar, N. R. Pal, B. Chanda and J. D. Sharma, “Detection of fraudulent alterations in ball point pen strokes using support vector machines”, In Proc. INDCON, 1-4, 2009.

[28] B. Halder and U. Garain, “Color feature based approach for determining Ink age in printed documents”, In Proc. ICPR, 3212-3215, 2010.

[29] J. V. Beusekom, F. Shafait and T. M. Breuel, “Text line examination for document forgery detection”, IJDAR, 189-207, 2013.

[30] R. D. S. Barboza. R. D. Lins and D. M. D. Jesus, “A color based model to determine the age of documents for forensic purpose”, In Proc. ICDAR, 1350-1353, 2013.

[31] R. Berrand, P. G. Kramer, O. R. Terrades, P. Franco and J. M. Ogier, “A system based on intrinsic features for fraudulent document detection”, In Proc. ICDAR, 106-110, 2013.

[32] A. Ahmed and F. Shafait, “Forgery detection based on intrinsic document features”, In Proc. DAS, 252-256, 2014.

[33] R. D. S. Barboza, R. D. Lins, E. D. F. D. Lira and A. C. A. Camara, “Later added strokes of text fraud detection in documents written with ballpoint pens”, In Proc. ICFHR, 517-522, 2014.

[34] Z. Luo, F. Shafait and A. Mian, “Localized forgery detection in hyperspectral document images”, In Proc. ICDAR, 496-500, 2015.

[35] Z. Khan, F. Shafait and A. Mian, “Automatic Ink mismatch detection for forensic document analysis”, PR, 3615-3626, 2015.

[36] K. S. Raghunandan, P. Shivakumara, B. J. Navya, G. Pooja, N. Prakash, G. Hemantha Kumar, U. Pal and T. Lu, “Fourier coefficients for fraud handwritten document classification through age analysis”, In Proc. ICFHR, 25-30, 2016.

[37] M. Mukthar and D. Malhotra, “$S_iS_bD_p$-The technique to identify forgery in legal handwritten documents”, In Proc. ICSSIT, pp 1103-1108, 2020.

[38] L. Nandanwar, P. Shivakumara, S. Kundu, U. Pal, T. Lu and D. Lopresti, “Chebyshev-Harmonic-Fourier-Moments and deep CNNs for detecting forged handwriting”, In Proc. ICPR, pp 6562-6569, 2021.

[39] L. Nandanwar, P. Shivakumara, U. Pal, T. Lu, D. Lopresti, B. Seraogi and B.B. Chaudhuri, “A new method for detecting altered text in document images”, International Journal of Pattern Recognition, Vo. 35, 2021.

[40] P. Ye and D. Doermann, “Document image quality assessment: A brief survey”, In Proc. ICDAR, 723-727, 2013.

[41] G. Liang, P. Shivakumara, T. Lu and C. L. Tan, “A New Wavelet-Laplacian Method for Arbitrarily-Oriented Character Segmentation in Video Text Lines”, In Proc. ICDAR, 391-396, 2015.

[42] X. Yi and M. Eramian, “LBP based segmentation of defocus blur”, IEEE Trans. IP, 1626-1638, 2016.

[43] G. Huang, Z. Liu, L. V. D. Maaten and K. Q. Weinberger, “Densely connected convolutional networks”, In Proc. CVPR, pp 2261-2269, 2017.

[44] N. Stamatopoulos, B. Gatos, G. Louloudis, U. Pal and A. Alaei, “ICDAR 2013 handwriting segmentation contest”, In Proc. ICDAR, 1402-1406, 2013.

[45] A. Alaei, U. Pal and P. Nagaghushan, “Dataset and ground truth for handwritten text in four different scripts”, IJPRAI, 2012.