

Hyperspectral Imaging for Ink Mismatch Detection

Aqsa Khan, Sara Sheikh, Nazish Iqbal

Department of Electrical Engineering

Institute of Space Technology (IST), Islamabad, 44000, Pakistan

Email: khanaqsa318@yahoo.com, sheikhsara22@yahoo.com, nazishiqbal81@gmail.com

Abstract— The discovery of ink mismatch provides important clues to pre-writing examiners to indicate whether a particular manuscript is written by a specific pen, or if a particular part of a note (e.g., signature) is written on a different ink compared to any other ink. In this paper, we show that the hyperspectral image (HSI) of the handwritten notes differs between visually similar inks. For this purpose, we have created the first hyperspectral data domain for a handwritten image in various blue and black inks, containing 33 visual reference bands. In unsupervised clustering technique, visual responses of inks fall into different groups to allow the separation of two different inks from the text of the questions. The same method when used in the RGB scan of these outputs fails to accurately distinguish the ink because it is very difficult to separate the ink from the optical range. HSI overcomes RGB deficiencies and allows better discrimination between inks.

Key Words—Hyperspectral images, K-Means clustering, ink mismatch detection.

I. INTRODUCTION

The human eye is exposed to light in the visible range and can distinguish different colors [1]. However, people tend to be consistent on certain colors, which mean they cannot distinguish between two colors that are close. But the details they see differ.

One of the most effective ways to correctly classify image pixels in an optical system is through image segmentation. In the areas of health care, image processing, traffic imagery, pattern recognition and more, image segmentation is very important (Ostu et al., 2015) [10]. The purpose of image segmentation by pixels within an image is to extract the most important data for extracting useful information. Examination of the ink is very important when analyzing the paper at issue. The result of an ink analysis can lead to fraud and ink age determination. One of the most important tasks is discriminating between the various inks. There are two key ways to make a distinction between inks, namely destructive and non-destructive tests. For the separation of parts of the ink mixture, chemical analysis such as thin layer chromatography (TLC) [2] was used. Ink separation is achieved by capillary action. This strategy raises some obstacles. First, testing is heat-sensitive, destructive, and time-consuming.

Alternatively, there is tremendous potential for a non-destructive testing process, as suggested by the visual system. Recently hyperspectral imaging has emerged as an important method for the identification, development [3], enhancement, comparison and prediction of forensic findings [4]. Such projects are important for investigators of a research paper by differentiating into writing ink. However, this task is difficult

and time consuming, as the tester must manually view the document under each wavelength and make a decision based on quality analysis. Depending on the perceived outcome of the thought process, the number of different levels of recognition can vary from a few to hundreds. Therefore, automated ink analysis plays an important role in supporting well documented document testing.

Brauns et al. [5] developed hyper spectral imaging to detect fraudulent documents. Their imaging system is based on an interferometer that relies on specific transmission frequency components and thus simplifies the detection process. Their notes serve as a proof of concept for finding literary references. A small sample shows that qualitative results in the dataset can be classified in different ways in an unsupervised manner.

More advanced hyperspectral imaging system developed for the analysis of historical documents in archives and libraries at the Netherlands National Archives [6]. The system provided high spatial and spectral resolution, visible by near-UV to the nearest IR range. The only drawback was that acquisition time was very slow [7]. Our proposed method, by comparison, collects hyperspectral images were just a part of the time.

Morales et al. [12] proposed a pen verification process, and Study of inks in handwritten documents using a square minimum Classification by SVM. Silva et al. [13] developed a destructive method to detect fraud on the basis of different chemo metrics in the documents. Khan et al. [14] proposed a joint scattering band selection-based hyperspectral imaging document analysis technique to separate different metameric veins. This work, however, uses an equal ink ratio to conduct tests and most non-proportionate cases have poor distinction and low precision, restricting their use in most practical cases.

M.J et al. [15] proposed the use of multispectral image analysis as an effective automatic ink mismatch detection method. Local entry separates the ink pixels and Fuzzy C-Means Clustering (FCM) is used to segment the ink pixel spectral response vectors into different clusters relative to the various inks used in the paper. In addition, the efficiency of feature selection in the clustering of ink pixels is also addressed.

A.et el. [16] proposed an ink analysis technique based on hyperspectral coordination proposed for the detection of ink mismatch. The main focus is to isolate the same transparent inks mixed in different ways to create the problem of uneven assembly.

A.U et al. [17] proposed a novel author-based approach to HSI and deep learning. Visual responses of text pixels and neighboring pixels through a document image are extracted and then transmitted to CNN for the author's distinction. Three different methods of visualization are used to use the atmosphere and visual information in handwritten notes. Reliable results are obtained after extensive research in the

publicly available database of UWA Writing Inks Hyperspectral Images (WIHSI). The results achieved in HSI images are compared with RGB-aligned images, which show the importance of HSI over color imaging in the author's presentation.

We present in this paper an efficient hyperspectral imaging system for recording the ink mismatch analysis. The research is based on the premise that the same inks display the same visual responses and vary in appearance of different inks.

We don't think the visual effects of inks represent the writing styles of different topics. Therefore, contrary to the works that classify manuscripts [8] or traces of ink deposition [9], our work only looks at the visual responses of inks to ink discrimination writing.

II. METHODOLOGY

The proposed algorithm contains K-means clustering which is an unsupervised learning method used when we have unsupervised data. The main goal of this algorithm is to separate data points from data set into distinct categories or groups. K-is trying to separate the data set into K-Clusters using the objective function.

III. PROPOSED ALGORITHM

The proposed algorithm is followed as below

1. Upload an image to split
2. Select the number of clusters K.
3. Choose from random K fields, centriods.
4. Assign each data point to the nearest centriods→ form a set of K clusters.
5. Calculate and set a new centroid for each cluster.
6. Re-assign each data point to the new closest centroid. If any recurrence occurs, go to step 4, if not, the model is ready.

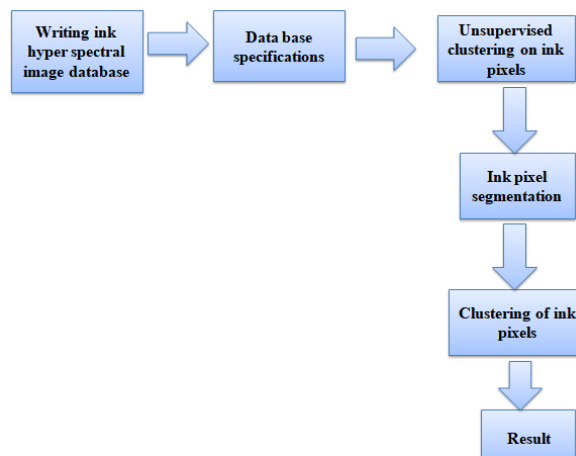
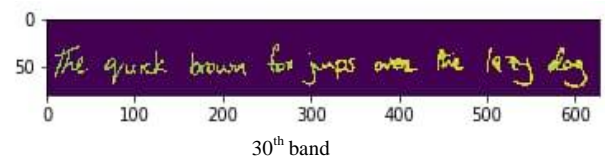
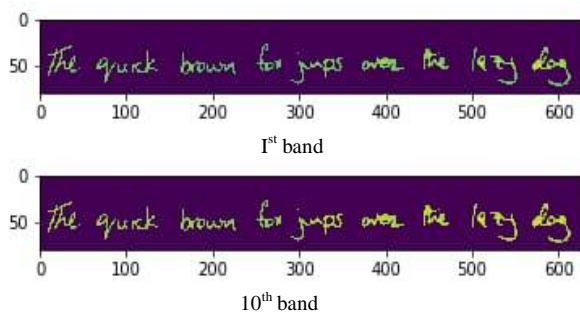
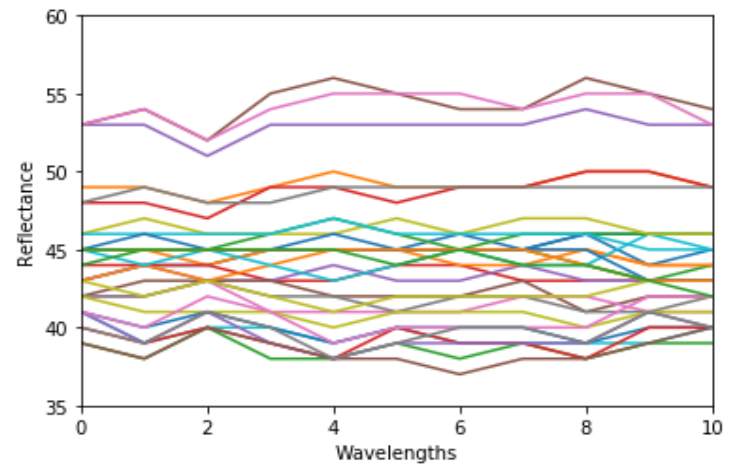


Figure 1 Block Diagram of Proposed Algorithm

IV. RESULTS



Compared to the RGB version of the images, the grayscale result shows us the usage of different inks at every band



The spectrum of the foreground of all the bands shows that there are different inks used rather than just one as there's a variation of the spectrum between 35-60.

We used two approaches here:

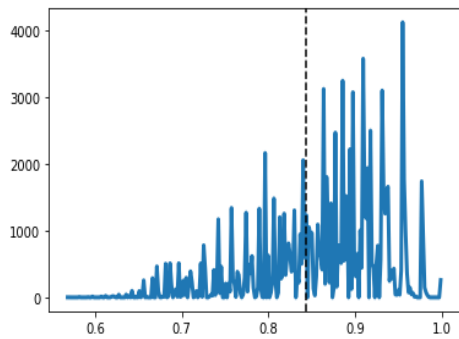
- K-mean clustering
- Elbow method

First we performed K-mean clustering with k=2 to the combined foreground of the bands. Following is the graph obtained of all the foregrounds and inks percentage in it:



Percentage of inkA in all bandds:32.759570627909184%
Percentage of inkB in all bands:67.24042937209082%

We then verified the thresholding point using Otsu thresholding technique [10] and almost same separation point. The histogram is as follows:

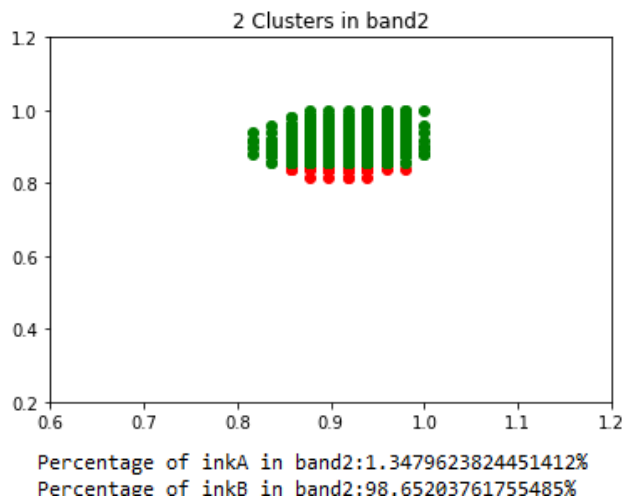


The thresholding point according to OTsu technique is 0.8436042

Then using the clusters formed we distinguished the percentage of inks in each band (here we illustrate clustering in band2, band12 & band25)

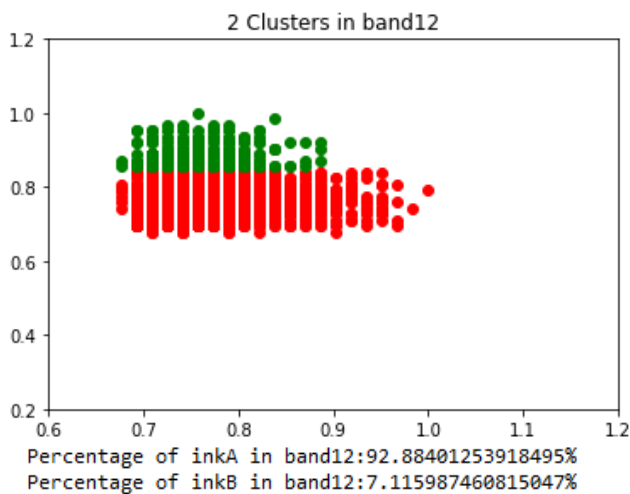
The quick brown fox jumps over the lazy dog

Band2



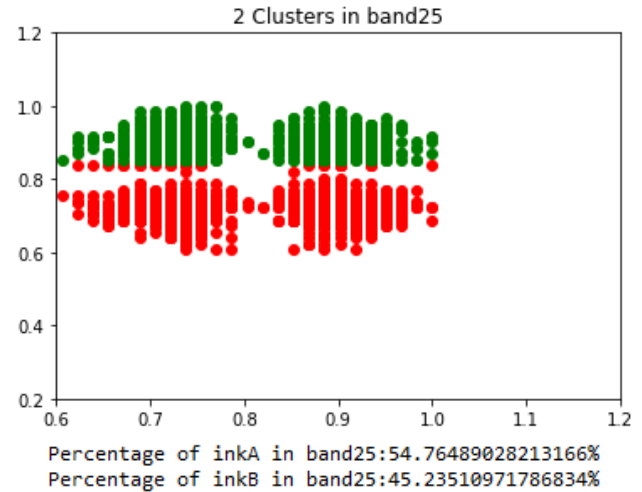
The quick brown fox jumps over the lazy dog

Band12



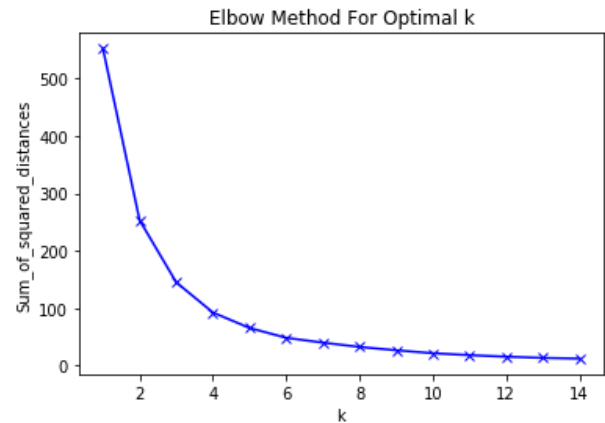
The quick brown fox jumps over the lazy dog

Band25



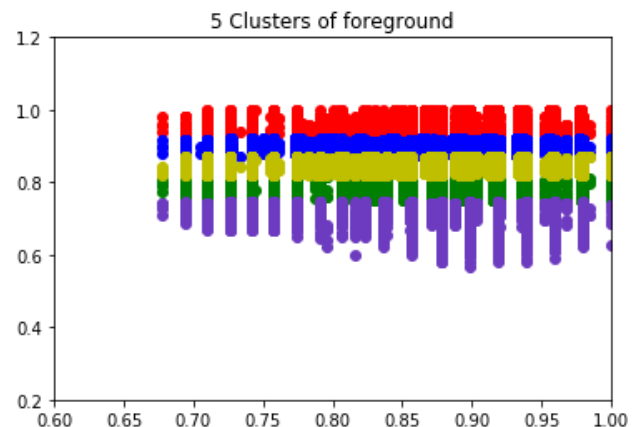
So by K-means we concluded that there are 2 major inks namely black and blue.

Second approach that we used was the elbow method [11] where we performed clustering on the foreground obtained by all bands and performed clustering from $k=1$ to $k=15$ and plotted it against the sum of squared distances for all k 's to get the optimal k . If the plot looks like an arm, then the elbow on the arm is optimal k . Following is the plot obtained:



Here in the plot above the elbow is at $k=5$ indicating the optimal k for this dataset is 5, so number of inks are 5

So after this we performed k mean clustering with $k=5$ on the foregrounds and we got the following graph:



We got following percentages of the inks in all bands:

Percentage of ink A:22.63%
 Percentage of ink B:14.08%
 Percentage of ink C:8.77%
 Percentage of ink D:31.38%
 Percentage of ink E:23.14

We then assigned colors to each cluster and print the bands.

Cluster1=red, Cluster2=blue, Cluster3=green, Cluster4=pink,
 Cluster5= yellow

Following is implementation on band1,band20, band30 &band33 with
 cluster1 as common in all :

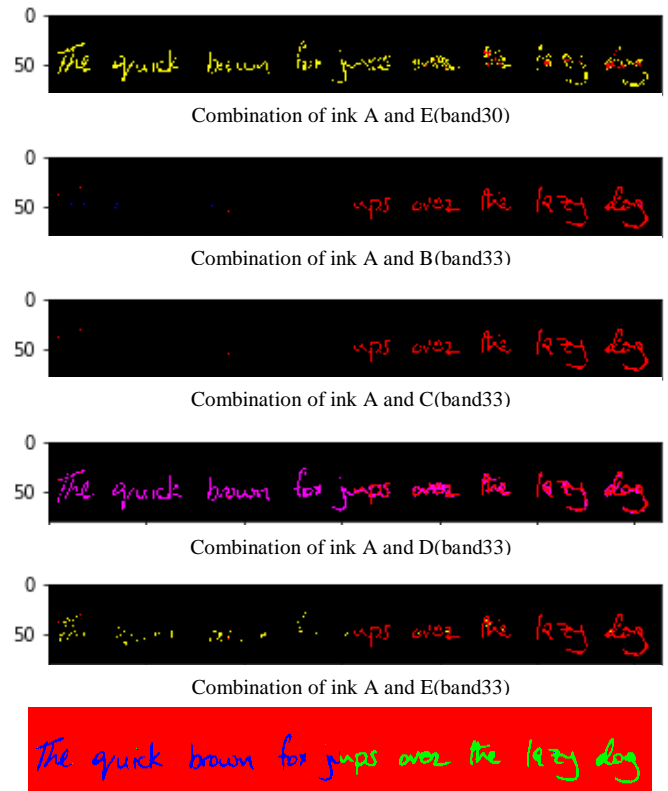
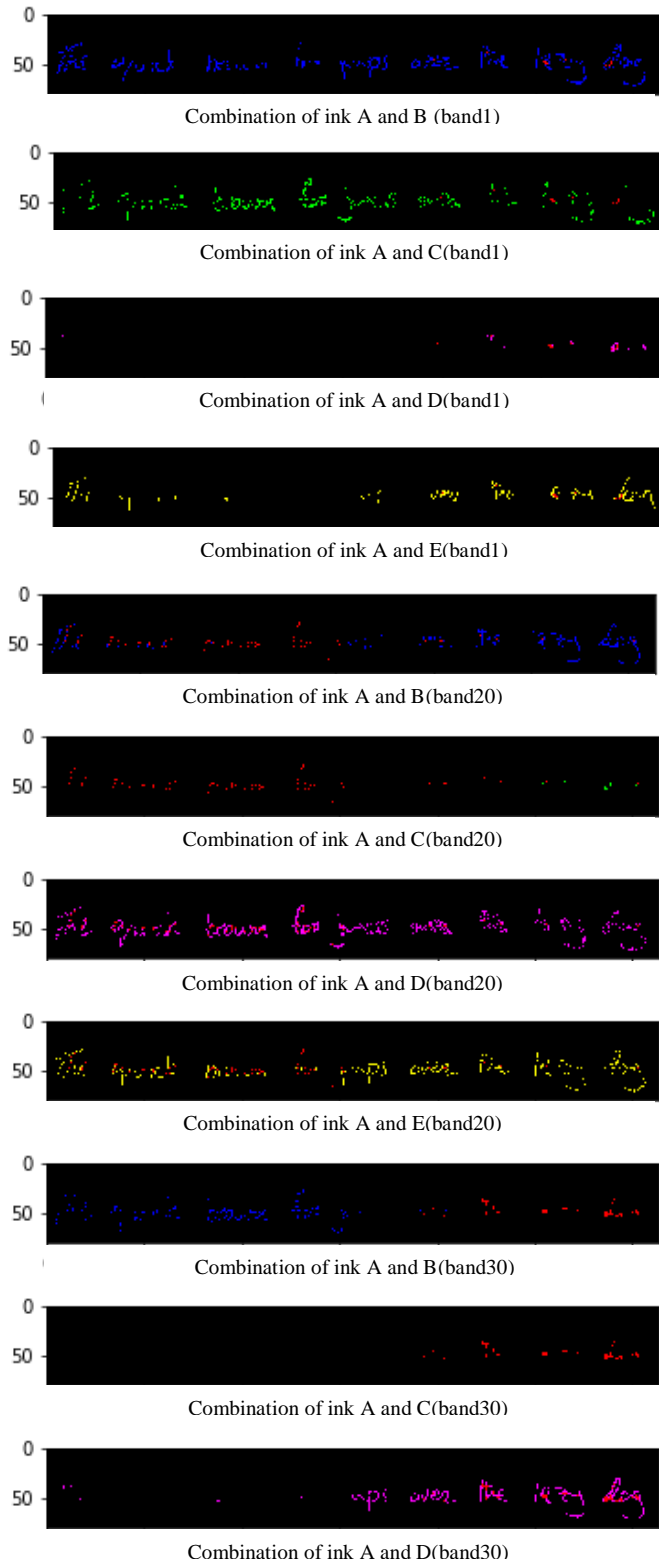


Figure 2 Output Image

Results show that both methods have great potential for identification of ink mismatch, as shown in Figure 2. We see that the bulk of the ink pixels are arranged correctly. Compared to the former methods, the new approach gives greater differentiation between inks in questioned documents.

V. CONCLUSION

Hyperspectral imaging is essential to support the testing of inks. We also gathered a collection of hyperspectral images of specific black and blue inks. We evaluated the output of RGB images and compared them to HSI images. We also investigated how the ink separation works with a collection of useful bands from hyperspectral images. Despite our hardware limitations regarding the signal-to - noise ratio, encouraging results have been achieved for ink separation.

REFERENCES

- [1]. E. H. Land, J. J. McCann et al., "Lightness and retinex theory," Journal of the Optical society of America, vol. 61, no. 1, pp. 1–11, 1971.
- [2]. V. Aginsky, "Forensic examination of "slightly soluble" ink pigments using thin-layer chromatography," Journal of Forensic Sciences, vol. 38, pp. 1131–1131, 1993.
- [3]. S. Joo Kim, F. Deng, and M. S. Brown, "Visual enhancement of old documents with hyperspectral imaging," Pattern Recognition, vol. 44, no. 7, pp. 1461–1469, 2011.
- [4]. G. Edelman, E. Gaston, T. van Leeuwen, P. Cullen, and M. Aalders, "Hyperspectral imaging for non-contact analysis of forensic traces," Forensic Science International, vol. 223, pp. 28–39, 2012.

- [5]. E. B. Brauns and R. B. Dyer, "Fourier transform hyperspectral visible imaging and the nondestructive analysis of potentially fraudulent documents," *Applied spectroscopy*, vol. 60, no. 8, pp. 833–840, 2006.
- [6]. R. Padoan, T. A. Steemers, M. Klein, B. Aalderink, and G. de Bruin, "Quantitative hyperspectral imaging of historical documents: technique and applications," *ART Proceedings*, 2008.
- [7]. M. E. Klein, B. J. Aalderink, R. Padoan, G. De Bruin, and T. A. Steemers, "Quantitative hyperspectral reflectance imaging," *Sensors*, vol. 8, no. 9, pp. 5576–5618, 2008.
- [8]. K. Franke, O. Bunnemeyer, and T. Sy, "Ink texture analysis for writer identification," in *Proc. IEEE Workshop on Frontiers in Handwriting Recognition*, 2002, pp. 268–273.
- [9]. K. Franke and S. Rose, "Ink-deposition model: The relation of writing and ink deposition processes," in *Proc. IEEE Workshop on Frontiers in Handwriting Recognition*, 2004, pp. 173–178.
- [10]. N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.
- [11]. Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, 1(6), 90-95.
- [12]. A. Morales, M. A. Ferrer, M. Diaz-Cabrera, C. Carmona, and G. L. Thomas, "The use of hyperspectral analysis for ink identification in handwritten documents," in *Security Technology (ICCST), 2014 International Carnahan Conference on. IEEE*, 2014, pp. 1–5.
- [13]. C. S. Silva, M. F. Pimentel, R. S. Honorato, C. Pasquini, J. M. Prats-Montalban, and A. Ferrer, "Near infrared hyperspectral imaging for forensic analysis of document forgery," *Analyst*, vol. 139, no. 20, pp. 5176–5184, 2014.
- [14]. Z. Khan, F. Shafait, and A. Mian, "Automatic ink mismatch detection for forensic document analysis," *Pattern Recognition*, 2015, (In Press).
- [15]. Khan, M. J., Yousaf, A., Khurshid, K., Abbas, A., & Shafait, F. (2018, April). Automated forgery detection in multispectral document images using fuzzy clustering. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)* (pp. 393-398). IEEE.
- [16]. Abbas, A., Khurshid, K., & Shafait, F. (2017, November). Towards automated ink mismatch detection in hyperspectral document images. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (Vol. 1, pp. 1229-1236). IEEE.
- [17]. Islam, A. U., Khan, M. J., Khurshid, K., & Shafait, F. (2019, December). Hyperspectral Image Analysis for Writer Identification using Deep Learning. In *2019 Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-7). IEEE.