International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

Analyzing Document Age with LBP Descriptors

Dr. Pushpalata Gonasagi

Associate Professor, Department of Computer Science, Govt. First Grade College Mahgaon Cross, Tq. Kamalapur, Dist. Kalaburagi, Karnataka, India

Abstract*— Paper documents have been the main source of information for communication, record-keeping and processing. A sheet of paper on which something is written or printed is referred to as a paper document. Paper documents are essential in contemporary society, enabling various aspects such as private letters and treaties. Paper documents necessitate security processes to ensure integrity and authenticity because these documents form the backbone of legal transactions. The security measures in documents include watermarking and embedded signatures, among other complex processes, which make forgery difficult. Functionalities not only secure significant legal documents but also foster trust among parties by ensuring their authenticity and trustworthiness. Paper documents play a crucial role in transitioning to a digital world, but digital technology has exposed security issues. One important stage in determining a document's originality is determining its age, which helps to solve this issue. This article uses 500 printed papers released between 1993 and 2013, with a 5-year gap, to determine whether the documents are new or old based on the year of publication. The proposed method was identified by identifying the year of publication of the paper document. Regardless of the text, line, logo, noise, etc., we have taken the entire document into consideration for analysis. First, we divided a page of a document into 512 x 512 blocks, keeping the text blocks that included the entire text. Here, a technique for determining a document's age based on Local Binary Pattern (LBP) features is provided, that subsequently used the LBP technique to extract characteristics from these text blocks. The K-Nearest Neighbors (KNN) classifier receives these features, and it achieves an average classification accuracy of new documents is 95.4% and 96.9% for old documents.

***Keywords*— Classification, KNN, LBP, Source, Printer, Segmentation.**

I. INTRODUCTION

In our everyday lives, paper records—such as identity cards, birth and death papers, official documents, etc.—are essential for maintaining records. However, as digital technology has advanced, there has been a growing concern regarding the legitimacy of these paper papers. Security issues persist despite the use of watermarking systems, printing patterns, logos, etc. during document preparation. In forensic science, classifying documents according to their age is a crucial responsibility.

A document's color changes as it becomes older. The document's texture and color are influenced by a number of factors, including temperature, humidity, storage conditions, surroundings, and frequency of access. It's possible that paper records were created centuries, decades, years, or even months ago.

In actuality, the technology used in the production, printing, and quality of paper have been evolving over time. Additionally, digital tools like Adobe Photoshop, Gimp, and Color Printer can be used maliciously to alter printed documents. Examining the document's uniqueness is crucial in this situation for a number of reasons. To assist the forensic science specialists in assessing the papers' legitimacy, we offer a technique for handling document age identification utilizing LBP. The remainder of the paper is structured as follows: The related work is described in section II, the proposed approach is presented in section III, the findings and discussion are in section IV, and the contributions of the proposed work are concluded in section V.

II. RELATED WORK

To support the originality of the inquiry papers, several studies have been conducted to categorize printed and handwritten documents as either new or ancient based on their age. Here, we have provided a concise overview of the research on document classification according to age. According to the papers' quality, Raghunandan et al. [1] described how to distinguish between old and new documents. They have employed Fourier co-efficient features in addition to foreground and background information from handwritten documents. They obtained an accuracy of 77.5% for old papers and 78.5% for fresh documents. He et al. [2] suggested utilizing scale-invariant traits to date historical documents. Historical documents' stroke shapes and evolutionary self-organizing maps are being used to determine how visual aspects have changed throughout time. The accuracy of this approach was 85.1%. A technique for estimating the dates of publication of printed historical texts was presented by Li et al. [3]. They employed Convolutional Neural Networks (CNNs) and Word Embedding to create a hybrid model that uses both text and visual input.

The result outperformed in both model's text and images. A technique that illustrates the application of hyper-spectral imaging was presented by Khan et al. [4]. for analyzing handwritten documents for fraud. Handwritten documents are found to have ink incompatibilities. However, the same handwritten paper was produced for illegal purposes using sophisticated software that matched the original documents' characters. A technique to ascertain the ink age of printed documents was put out by Biswajit et al. [5]. The foundation of this method is color picture analysis, which determines the image's kurtosis, pixel profile, and average intensities. These characteristics are used to determine the unknown samples' ink age. The dataset consisted of the cover pages of Google Life magazine. The recognition accuracy of a neural network, which is intended for ink age identification, was 74.5%. Using the RGB color components of the document's background, da Silva Barboza et al. [6] determine the document's age. The 20th century birth, wedding, and other certificates were taken into consideration for testing. They experimented with the document images, which were separated by fifty-two years, and obtained result outperform. A method that distinguishes the document from a collection of documents printed by various printers was introduced by Gebhardt et al. [7]. The impact of printing the characters varies depending on the printer. They have observed variations in the printed character edges between inkjet and laser printers. They classified the documents produced by laser and inkjet printers using the global KNN algorithm and the unsupervised anomaly Grubb's test. Bertrand et al. [8] presented a technique that automatically identifies character forgeries. The copy and paste methods that are frequently employed to create fake documents have been uncovered. By comparing the characters' similarity, the copy and paste characters are identified. They looked for odd character shapes to identify the simulation fraud. They obtained findings of 0.82 precision and 0.77 recall.

III. Proposed Method

Three crucial processes carried out by this method are the pre-processing, feature extraction, and classification of the documents. Fig.1 shows the block diagram for the suggested approach.

A. Pre-processing

Our collection consists of 100 document pages from five different publications published in 1993, 1998, 2003, 2008, and 2013.

The HP flatbed LaserJetProM128fn scanner, which has a resolution of 300 dpi, is used to scan these papers. After converting these scanned photos to grayscale, they are divided into text blocks measuring 512 by 512 pixels, with only the blocks completely covered by the text kept. Table 1 displays the dataset's specifics. As compared to reference [11], this proposed method increases the dataset and analyses the efficacy of the classification documents.

B. Feature Extraction

LBP: It is one of the efficient texture descriptors. LBP [9, 10] measures local spatial patterns and gray scale contrast of the underlying texture. It is robust for computing monotonic gray-scale changes in the image. Rotation invariant is the basic property of the LBP. Thus, we choose LBP to obtain proficient texture descriptors of text blocks. In number of applications, it has been proved that LBP is more robust than statistical texture features. Mathematically, LBP can be expressed as shown in Eq. 1. Where, P indicates sampling points on a circle of radius R (P = 8, R = 1), gc corresponds to the gray value of the center pixel and gp corresponds to the gray values of its neighbor pixel p [11].

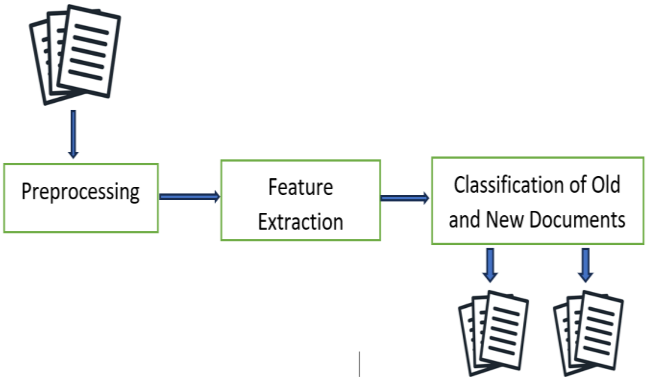

Fig. 1 Flow diagram of proposed method

Table 1
Details of the Dataset Developed for Experiments

Year	No. of doc images	No. of text blocks extracted
2013	100	1000
2008	100	1000
2003	100	1000
1998	100	1000
1993	100	1000
Total	500	5000

$$\mathrm{LBP}_{P,R} = \sum_{p=0}^{p-1} S(g_p - g_c)2^p \text{ where } S(x) \begin{cases} 1 & \text{if } x \geq 1 \\ 0 & \text{if } x < 0 \end{cases} \qquad (1)$$

C. *Classification*

Although we have used K-nearest neighbor (KNN) to determine the age of documents, it is a straightforward and naive classifier. Because LBP features are more discriminative in nature, we selected it over complex classifiers.

IV. Results and Discussion

Those printed in 2008, 2003, 1998, and 1993 are regarded as old for the sake of this experiment, while those printed in 2013 are regarded as new. In order to calculate the distance between the feature vectors of documents printed in 2013 and those printed in 2008, 2003, 1998, and 1993, respectively, we created two class issues and used KNN. Table 2 displays the categorization accuracy. Interesting findings from Table 2's classification accuracy indicate that accuracy rises in tandem with the documents' age gap. The reason for this is that documents printed in 1993 are twenty years older than those printed in 2013. Thus, 98.5% is the categorization accuracy shown in Table 2. Between a similar way, there is a ten-year age difference between documents printed between 2003 and 2013. As a result, there is greater resemblance between the documents from 2013 and 2008; thus, there is less accuracy, which stands at 93.2%. Table 3 displays the. Classification accuracy by considering documents of 2013 as new and others as old. The comparison of the results with the published work is not significant. Since the experimental context (different dataset, method, and number of features) is not identical. Furthermore, no age identification work has been documented on Kannada script documents. However, Table 4 presents a numerical comparison analysis, with the work referenced at [1, 11]. Age identification of handwritten documents with varying age gaps is the subject of this investigation.

Table 2:
Classification accuracy by considering documents of 2013 as new and others as old

Year (new/old document)	2008	2003	1998	1993
2013	93.2%	94.9%	97.6%	98.5%

Table 3:
Classification accuracy by considering documents of 2013 as new and others as old

Accuracy	New documents (2013)	Old documents (1993–2008)
New documents (2013)	95.4%	4.6%
Old documents (1993–2008)	3.1%	96.9%

Table 4:
Classification rate of the proposed system without block-wise dataset compared to existing approach

Method	Identification of New documents	Identification of Old documents
Raghunandan et al. [1]	78.5%	77.5%
Gonasagi et al. [11]	87.5%	95%
Proposed Method	95.4%	96.9%

V. Conclusion

In this study, we suggested a new method for dividing document images into new and old categories. When it comes to differentiating between new and old text blocks, the LBP has demonstrated impressive results. LBP produced high-discriminating features and effectively recorded the documents' textural attributes. Document age identification research is carried out to investigate the nature of the documents to certify its originality. In the future, we are expanding it to determine the handwritten documents' age.

REFERENCES

[1] Raghunandan, K.S., B.J. Palaiahnakote Shivakumara, B.J. Navya, G. Pooja, Navya Prakash, G. Hemantha Kumar, Umapada Pal, and Tong Lu. 2016. Fourier coefficients for fraud handwritten document classification through age analysis. In 2016 15th International conference on frontiers in handwriting recognition (ICFHR), 25–30. IEEE.

[2] He, Sheng, Petros Samara, Jan Burgers, and Lambert Schomaker. 2016. Discovering visual element evolutions for historical document dating. In 2016 15th International conference on frontiers in handwriting recognition (ICFHR), 7–12. IEEE.

[3] Li, Yuanpeng, Dmitriy Genzel, Yasuhisa Fujii, and Ashok C. Popat. 2015. Publication date es- timation for printed historical documents using convolutional neural networks. In Proceedings of the 3rd international workshop on historical document imaging and processing, 99–106. ACM.

[4] Khan, Zohaib, Faisal Shafait, and Ajmal Mian. 2015. Automatic ink mismatch detection for forensic document analysis. Pattern Recognition 48 (11): 3615–3626.

[5] Halder, Biswajit, and Utpal Garain. 2010. Color feature based approach for determining ink age in printed documents. In 2010 20th International conference on pattern recognition, 3212-3215. IEEE.

[6] da Silva Barboza, Ricardo, Rafael Dueire Lins, and Darlisson Marinho de Jesus. 2013. A color-based model to determine the age of documents for forensic purposes. In 2013 12th International conference on document analysis and recognition, 1350–1354. IEEE.

[7] Gebhardt, Johann, Markus Goldstein, Faisal Shafait, and Andreas Dengel. 2013. Document authentication using printing technique features and unsupervised anomaly detection. In 2013 12th International conference on document analysis and recognition, 479–483. IEEE.

[8] Bertrand, Romain, Petra Gomez-Krämer, Oriol Ramos Terrades, Patrick Franco, and Jean- Marc Ogier. 2013. A system based on intrinsic features for fraudulent document detection. In 2013 12th International conference on document analysis and recognition, 106–110. IEEE.

[9] Ojala, Timo, Matti Pietikäinen, and David Harwood. 1996. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition 29 (1): 51– 59.

[10] Ojala, Timo, and Matti Pietikäinen. 1999. Unsupervised texture segmentation using feature distributions. Pattern Recognition 32 (3): 477–486.

[11] Gonasagi, Pushpalata, Rajmohan Pardeshi, and Mallikarjun Hangarge. "Classification of documents based on local binary pattern features through age analysis." In Ambient Communications and Computer Systems: RACCCS 2019, pp. 265-271. Singapore: Springer Singapore, 2020.