Research Article

Intelligent Bar Chart Plagiarism Detection in Documents

**Mohammed Mumtaz Al-Dabbagh,^{1,2} Naomie Salim,¹
Amjad Rehman,³ Mohammed Hazim Alkawaz,^{1,2} Tanzila Saba,⁴
Mznah Al-Rodhaan,⁵ and Abdullah Al-Dhelaan⁵**

¹ Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

² Faculty of Computer Sciences and Mathematics, University of Mosul, Mosul, Iraq

³ MIS Department, CBA, Salman Bin Abdulaziz University, Alkharj, Saudi Arabia

⁴ College of Computer and Information Sciences (CCIS), Prince Sultan University, Riyadh, Saudi Arabia

⁵ Computer Science Department, College of Computer & Information Sciences, King Saud University, Riyadh, Saudi Arabia

Correspondence should be addressed to Amjad Rehman; ar.khan@sau.edu.sa

Received 30 March 2014; Revised 21 June 2014; Accepted 7 July 2014; Published 17 September 2014

Academic Editor: Iftikhar Ahmad

Copyright © 2014 Mohammed Mumtaz Al-Dabbagh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a novel features mining approach from documents that could not be mined via optical character recognition (OCR). By identifying the intimate relationship between the text and graphical components, the proposed technique pulls out the Start, End, and Exact values for each bar. Furthermore, the word 2-gram and Euclidean distance methods are used to accurately detect and determine plagiarism in bar charts.

1. Introduction

Detection, determination, and rectification of plagiarism are outstanding quests in every sphere of documentation and copyright. Lately, the significant advancement in information technology represented by digital libraries and World Wide Web is regarded as one of the main reasons for exponential growth in plagiarism appearance. It has become effortless for the plagiarist to utilize or copy the work of others without acknowledging or citing them due to the easy availability of most resources in digital format. Thus, plagiarism is regarded as one of the electronic crimes and intellectual thefts from others documents [1–3]. In academia, plagiarism posed a severe educational challenge which is acutely faced by research institutions, universities, and even schools. Several efforts are dedicated to detecting different types of plagiarism via programming code and text. Plagiarism detection began in the 1970s, where the identification of rate of plagiarism in programming code written by some computer languages such as C and Pascal was introduced [4]. Digital documents

being the major carriers of information require extreme authentication in terms of their origins and trustfulness. The quest for achieving an accurate and efficient image forgery detection method in digital documentation is never ending. Developing a robust plagiarism detector by overcoming the limitations associated with human intervention is the key issue [5].

Recently, several researchers developed the algorithmic approach using computer codes to detect plagiarism in the homework of students [6]. Based on levels of plagiarism patterns some studies introduced plagiarism detection methods which are implemented in the algorithms and program codes [7]. Generally, the computerized or statistical approaches are exploited to detect plagiarism in natural language since the 1990s. The techniques used for natural language are based on various factors such as grammar, semantic, and grammar-semantics hybridizations [1, 4]. However, the grammar-based method is one of the restrictive ones to detect plagiarism. This type of method analyzes the sentences based on grammatical structure, which can be efficiently used to detect the Exact

FIGURE 1: Taxonomy of chart plagiarism.

Copy of text. While semantic-based method utilizes vector space model to calculate the similarities among the texts. Undoubtedly, the grammar-semantics hybrid approach overcomes all disadvantages of the other methods. This is considered as one of the most versatile techniques to detect text plagiarism [1, 8, 9].

A new taxonomy is introduced to explain the concepts for various types and patterns of text plagiarism [4]. Plagiarism is divided into two main parts including literal and intelligent one. Each part consists of several subparts which cover all possibilities of text plagiarism. Generally, the representations of quantitative information are formulated via infographic form by using figures, charts, and tables. The information that is displayed in charts, figures, and tables includes results of experiments, framework, and statistical facts. These data and information in homogenous form can be formulated by using various shapes such as pie chart, bar chart, and 2D and 3D plots [10–12].

We report a new type of plagiarism detection method by highlighting the types of information that can be stolen from others work without referencing. Firstly, different types of forged information are organized into taxonomy of chart,

figure, and table to highlight varieties of plagiarism patterns such as Exact and Modified Copy. Secondly, plagiarism detection in bar chart image is performed depending on ten features in images. Some of the features are extracted by OCR tool while others are acquired from the relationship of text and graphic components [13, 14]. Finally, the proposed technique is used to extract the features of bar chart images which cannot be extracted by OCR to detect plagiarism. The paper is organized as follows: Section 2 describes various existing techniques for extracting data from bar chart images [15]. The taxonomy of chart, figure, and table related to plagiarism is presented in Section 3. Section 4 discusses the methodology and Section 5 includes the experimental results of bar chart plagiarism detection. The discussions are elucidated in Section 6. Section 7 concludes the paper.

2. Related Work

Categorizations of bar chart images refer to their labeling into one of the predefined geometrical or nongeometrical classes. Though the classification is apparently manageable,

FIGURE 2: Taxonomy of figure plagiarism.

it is proven to be an extremely difficult problem in computer programming. Hence, there is an intense attention in developing automatic tools to categorize, describe, or retrieve images based on their contents.

Consequently, researchers attempted to extract the features and data from chart images. For automatic images categorization and description computational model is successfully introduced [16]. The analysis of local and global image characteristics by using text and image features is used in the model. The model is capable of differentiating geometrical and ordinary images. The computational model is comprised of classifier stage which is trained by the associated text features using advanced concepts and similarity matching stage.

Classification methods based on multiple-instance learning for chart images are also developed [10]. A re-revision

system consisting of three concatenated major stages such as classification, extraction, and redesigned chart images is employed [17]. In the extraction stage, two types of charts (pie and bar) are focused on. Some techniques are presented in extracting data and graphical marks from chart images. Truly, the understanding and recognition of chart images require the preprocessing and extraction of data and information. Primarily, two types of available methods that deal with chart images are either to consider electronic chart directly [18–20] or to obtain them after converting into raster images [21–24]. Mishchenko and Vassilieva [25] introduced a model-based method for the classification of chart images which involved two main stages: firstly, predicting the location and the size of chart depending on the color distribution of chart image and secondly the extraction and matching of chart image edges to achieve the best match between query and database images.

FIGURE 3: Taxonomy of table plagiarism.

The techniques for features extraction of image depend on the type of images such as chart or medical representation. Some techniques are applicable on two-dimensional plot of chart images while others work well for bar chart images. Hough transform technique is introduced as an approach to extract the features of bar chart images [23]. Some investigations are based on the edges of bars to extract the features [24, 26]. The learning-based method is established to recognize the chart images [22]. The features of bar chart images can be extracted by describing the height and width of each bar, which is applied on statistical images to determine the similarities [27]. Meanwhile, other techniques focused on geometric features rather than data and information of scientific bar chart images [28].

Currently, several techniques are developed to extract the features from medical images. The texture is one of the visual

contents of a medical image used in content-based image retrieval (CBIR) to represent the image effectively for searching and recovering similar areas [29]. Gray-level statistical matrix technique is applied to extract the texture information for the content-based retrieval of mammograms from the MIAS database [30]. 3D texture features technique based on the cooccurrence matrixes of the gray-level, gradient, and curvature information regarding the nodule volume data for classifying the malignancy from benign is introduced [31, 32].

3. Plagiarism Taxonomy and Patterns

Three types of graphic plagiarism such as figure, chart, and table are important to emphasize. Each type highlights different levels of plagiarism. The patterns and types of plagiarism for figures, charts, and tables are presented as

taxonomy. Some kinds of text plagiarism are also evaluated [33]. The methods for detecting passages of text plagiarism for documents without appropriate citations are also suggested. Taxonomy is further extended to cover other types of plagiarism [4]. The taxonomy presented in various studies majorly demonstrates literal and intelligent plagiarism, where each kind includes many patterns of plagiarism. However, we are interested in detecting plagiarism of charts, as well as their taxonomy, figures, and tables. Alternatively, charts (pie, bar, and line) can be considered as one of the methods for representing the data and information of experimental results or comparing among techniques which are copied from other references without citation. Therefore, plagiarism of charts can be formulated in several forms to manifest the same information in various shapes. Taxonomy of chart plagiarism demonstrates many patterns and models which may be used to plagiarize the data of chart image.

Plagiarism patterns of chart, figure, and table are divided into Exact Copy and Modified Copy prototypes. The Exact Copy patterns of plagiarism are defined as the direct quote of data from other works without referencing, where copy and paste of the whole or part of the information image is performed. Simplicity is one of the important attributes of this type of plagiarism. Besides, this type of plagiarism does not require much time to hide the academic crime. The other type of graphic plagiarism is the Modified Copy for information of chart, figure, and table. This is more intelligently performed than the previous one because the same data can be formulated in many ways to exhibit the work in a different style than the original one. The goal of these intelligent means is that the plagiarist attempts to deceive the readers by doing some changes, such as translation from other languages or generating another shape for the same data.

The Modified Copy plagiarisms are primarily divided into translation and restructuring. In this research, new types of copying are organized by taxonomy which explains various patterns of graphic plagiarism. Furthermore, the primary focus of the bar chart image is to detect the proportion of plagiarism. Figures 1, 2, and 3 depict the taxonomy of chart, figure, and table plagiarism, respectively.

4. Methodology

The methodology of bar chart plagiarism detection as shown in Figure 4 consists of three main stages, namely, planning and collection, feature extraction, and development with system evaluation. In the planning and collection stage, various patterns of graphical plagiarism via taxonomy of chart, figures, and table are presented (Figures 1, 2, and 3). The taxonomy of chart plagiarism explains different formulations which plagiarize the data of bar chart images. Therefore, varieties of bar chart images are collected and data sets are considered. The gatherings of the data sets consist of 100 bar chart images for storing in databases and twenty images for query including all possibilities of plagiarism for bar chart images. These data sets are collected from different resources such as thesis, which represented various types of bar chart images in 2D and 3D. Besides, vertical and horizontal bar chart images as shown in Figure 5 are taken into account.

FIGURE 4: Flowchart of methodology.

In the feature extraction stage, the features of bar chart images are used to detect plagiarism. Various types of bar chart images are analyzed to detect the features of image. The bar chart images inferred to acquire maximum of ten features representing the information and the data of image. These features are common in different types of bar chart images, for instance, in 2D and 3D images. However, the number of uses for these features may differ from each other.

The features extraction is an essential process to get the data from images which can be utilized to detect the rate of plagiarized data. Therefore, these ten features are categorized into low-level and high-level features. The low-level features refer to the text features of bar chart. The text features are the text which can be used in image to represent the information and data such as caption of image, label of each bar, label of coordinates, and values on coordinates. Generally, they are extracted from bar chart images using OCR tool. Conversely, the high-level features referring to numeric features cannot be extracted using OCR tool. The extraction of numeric features requires a relationship between the text and graphic components. The numeric features include values of bars in image. Each bar in an image has three numeric features which can be extracted by the proposed technique depending on Start, End, and Exact values. The Start and End values represent the first and last values while the Exact one corresponds to the real value of the bar. For instance, the text features for image in Figure 5 are "Figure 10: Revision Design Galleries," "Silver, Platinum, ..., Cash," and "0%, 5%, ..., 40%" which represent caption of image and label of coordinates X and Y, respectively. Meanwhile, the numeric features represent the value of each bar, which can be detected by Start, End, and Exact features. For example, the Start, End, and Exact features of bar Silver are 5%, 10%, and 6%, respectively.

These features are used to detect the proportion of plagiarism for bar chart image. The extraction of Start, End, and

(a)

(b)

FIGURE 5: Samples of data set [17, 36].

Exact values necessitates preprocessing of bar chart image to the adjacent coordinates of the image. Image scanning is then performed to detect the length of each bar in order to find the numeric features for each bar. Storing of the features in databases depends on the type of features whether numeric or text. The features which are extracted by the proposed technique are represented as vectors, while the text features that are extracted by OCR are characterized as string.

The detection methods for text plagiarism are mainly categorized based on character, semantics, structure, citation, cluster, cross language, and syntax. Comparatively, the smaller number of textual components than normal paragraph text allows us to use character-based methods to detect plagiarism of bar chart images. The character-based methods depend on character matching approaches to exactly or partially detect the identical string for features of bar chart images. Various algorithms of plagiarism are adopted in the text as character n -gram to identify the similarity between two strings based on the number of identical characters of features. Some researchers use 8-gram and 5-gram techniques [34, 35] for matching strings to detect plagiarism. We used 2-gram technique to detect plagiarism of bar chart images. This technique is used to represent the text features of bar chart images. Different similarity measures can be used to obtain the similarity for numeric features such as Euclidean distance, Jaccard, or cosine coefficient. The Euclidean distance is calculated by the following:

$$Ec(x, y) = \sqrt{\sum_i |x_i - y_i|^2}. \quad (1)$$

Once the detection and storing of the proportion of plagiarism are completed then the performance of the system is evaluated. The performance is evaluated by overlapping of

features using the relation of Precision and Recall given by the following:

$$\begin{aligned} \text{Recall} &= \frac{\text{Relevant Documents Retrieved}}{\text{All Relevant Documents}}, \\ \text{Precision} &= \frac{\text{Relevant Documents Retrieved}}{\text{All Documents Retrieved}}. \end{aligned} \quad (2)$$

5. Experimental Results

The bar chart plagiarism detections are carried out in four main stages such as submission of query images, feature extraction of bar chart image, plagiarism detection, and highlighting results. The first stage is to submit various types of query images covering different kinds of possible plagiarism to detect and judge plagiarism of bar chart images, while the features of query are extracted in the second stage. The third stage includes detection of plagiarism by using word 2-gram and Euclidean distance techniques. Finally, the features of query bar chart image that are plagiarized from others are highlighted and the proportion of the similarity is displayed.

The bar chart plagiarism is further divided into *Exact Copy* and *Modified Copy* as explained in taxonomy. Figure 6(a) shows the query image while Figure 6(b) depicts the plagiarized images detected by the system. The first plagiarized image is similar to the whole data in the image while the second plagiarized image contains the same data that was plagiarized but presented as a horizontal bar chart image. The system extracts the features of query image and detects the proportion of plagiarism depending on Start, End, and Exact values for each bar as well as the label of each bar. The system highlights the data and information that are plagiarized and provides the proportion of plagiarism.

One of the patterns of plagiarism derived from *Modified Copy* is the stealing by changing scales. Each bar is modified

FIGURE 6: Plagiarism detection for one of Exact Copy patterns as plagiarism of the whole data of image.

by plagiarists by changing Start and End values to be different from the original image. Figure 7 illustrates the significant role played by the Exact values to detect this type of plagiarism.

The plagiarists may use integration among patterns of possible bar chart plagiarism to present a more complex image which has the same data quoting from other works. Figure 8 displays the query image which is modified by changing colours and scales of bars as well as changing their location via swapping. The proposed system is capable of detecting this type of plagiarism and identifies the proportion of similarity.

Figure 9 illustrates the performance of the system for plagiarism detection of *Exact Copy* and *Modified Copy* patterns, respectively.

6. Discussion

The state-of-the-art graphical plagiarism techniques and patterns are presented. The graphical plagiarism is considered one of the electronic crimes and thefts and the concepts of such stealing are newly viewed. Various important information and data can be represented as graphical forms such as results or frameworks for academic and business aspects. However, many systems of text plagiarism methods such as Turnitin are incapable of detecting plagiarism of images. In spite of the different styles of bar chart images, the extraction of features of image plays an important role in detecting plagiarism. Our proposed technique which is used to extract the numeric features played an essential role for bar chart plagiarism detection. The patterns of *Exact Copy* of bar chart

FIGURE 7: Plagiarism detection for one of the Modified Copy patterns which is the stealing by changing scales of image.

plagiarism detection including direct copy of all data or part of data are underscored. Plagiarism which is carried out by modifying caption of images via restructuring or summarizing for label sentences is emphasized. Alternatively, the patterns of *Modified Copy* which are regarded as more complicated than *Exact Copy* patterns are also analyzed. The difficulty of these patterns is the changing on image which appears as the same data and information in different forms. The restructuring of information for image within the same shape is also covered. The edition of bar chart images including the change of image bar colors or changing the bar locations either by swapping or via generating horizontal bars from vertical bars and vice versa is discussed in detail.

Besides, more professional modification such as changing of scales on coordinates which is completely different from original one can be detected by the proposed method. In this case, the *Start* and *End* features of bars are completely different. Consequently, the *Exact* features as well as other attributes play significant role in detecting plagiarism in bar chart image.

7. Conclusion

We demonstrate the precise recognition of different plagiarized patterns in business documents using an intelligent bar chart detection system. The types and patterns of plagiarism

FIGURE 8: Plagiarism detection for integration among possible bar chart plagiarism.

FIGURE 9: The evaluation of performance (a) for *Exact Copy* patterns and (b) for *Modified Copy* patterns.

are presented via taxonomy of figure, chart, and table. Various kinds of possible plagiarism are highlighted using taxonomy. Plagiarism of bar chart image as type of chart is detected by the newly proposed technique. It is established that the present technique is capable of extracting the features from a bar chart image which cannot be pulled out using OCR tool. Our technique first recognizes the connection between the text and graphical components to extract the Start, End, and Exact value for each bar. Using word 2-gram and Euclidean distance methods the accurate detection of plagiarism is performed. The detection of plagiarism is based on ten striking features. The system is capable of detecting different levels of plagiarism not only copy and paste of bar chart image but also modification on images such as changing color or scales. The present system efficiently and accurately distinguishes other possible alteration administered on these images such as swapping among bars location and even changes on caption via summarizing and restructuring. The proposed technique may be useful for intelligent plagiarism detection in business and academic documents.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through Research Group no. RGP-264. The authors are also thankful to Ministry of Science and Technology Innovation (MOSTI), Malaysia, and Research Management Center (RMC), Universiti Teknologi Malaysia (UTM), Johor, Malaysia, for their technical support and expertise in conducting this research.

References

- [1] A. M. El Tahir Ali, H. M. D. Abdulla, and V. Snasel, "Survey of plagiarism detection methods," in *Proceedings of the 5th Asia Modelling Symposium (AMS '05)*, pp. 39–42, May 2011.
- [2] A. Rehman and T. Saba, "Features extraction for soccer video semantic analysis: current achievements and remaining issues," *Artificial Intelligence Review*, vol. 41, no. 3, pp. 451–461, 2014.
- [3] A. Rehman and T. Saba, "Evaluation of artificial intelligent techniques to secure information in enterprises," *Artificial Intelligence Review*, pp. 1–16, 2012.
- [4] S. M. Alzahrani, N. Salim, and A. Abraham, "Understanding plagiarism linguistic patterns, textual features, and detection methods," *IEEE Transactions on Systems, Man and Cybernetics C: Applications and Reviews*, vol. 42, no. 2, pp. 133–149, 2012.
- [5] T. Saba and A. Altameem, "Analysis of vision based systems to detect real time goal events in soccer videos," *Applied Artificial Intelligence*, vol. 27, no. 7, pp. 656–667, 2013.
- [6] K. J. Ottenstein, "An algorithmic approach to the detection and prevention of plagiarism," *SIGCSE Bulletin*, vol. 8, pp. 30–41, 1976.
- [7] A. Parker and J. O. Hamblen, "Computer algorithms for plagiarism detection," *IEEE Transactions on Education*, vol. 32, pp. 94–99, 1989.
- [8] J. P. Bao, J. Y. Shen, X. D. Liu, and Q. B. Song, "Survey on natural language text copy detection," *Journal of Software*, vol. 14, no. 10, pp. 1753–1760, 2003.
- [9] T. Saba and A. Rehman, *Machine Learning and Script Recognition*, Lambert Academic Publisher, 2012.
- [10] W. Huang, S. Zong, and C. L. Tan, "Chart image classification using multiple-instance learning," in *Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV '07)*, p. 27, Austin, Tex, USA, February 2007.
- [11] T. Saba, A. Rehman, and M. Elarbi-Boudihir, "Methods and strategies on off-line cursive touched characters segmentation: a directional review," *Artificial Intelligence Review*, 2011.
- [12] A. Rehman, D. Mohammad, G. Sulong, and T. Saba, "Simple and effective techniques for core-region detection and slant correction in offline script recognition," in *Proceedings of the IEEE International Conference on Signal and Image Processing Applications (ICSIPA '09)*, pp. 15–20, Kuala Lumpur, Malaysia, November 2009.
- [13] A. Rehman and T. Saba, "Document skew estimation and correction: analysis of techniques, common problems and possible solutions," *Applied Artificial Intelligence*, vol. 25, no. 9, pp. 769–787, 2011.
- [14] A. Rehman and T. Saba, "Off-line cursive script recognition: current advances, comparisons and remaining problems," *Artificial Intelligence Review*, vol. 37, no. 4, pp. 261–288, 2012.
- [15] T. Saba, A. Rehman, and G. Sulong, "Improved statistical features for cursive character recognition," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 9, pp. 5211–5224, 2011.
- [16] T. Helmy, "A computational model for context-based image categorization and description," *International Journal of Image and Graphics*, vol. 12, no. 1, Article ID 1250001, 2012.
- [17] M. Savva, N. Kong, A. Chhajta, F.-F. Li, M. Agrawala, and J. Heer, "ReVision: automated classification, analysis and redesign of chart images," in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*, pp. 393–402, Santa Barbara, Calif, USA, October 2011.
- [18] S. Elzer, S. Carberry, I. Zukerman, D. Chester, N. Green, and S. Demir, "A probabilistic framework for recognizing intention in information graphics," in *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI '05)*, pp. 1042–1047, Edinburgh, UK, August 2005.
- [19] S. Carberry, S. Elzer, N. Green, K. McCoy, and D. Chester, "Understanding information graphics: a discourse-level problem," in *Proceedings of the 4th SIGDIAL Workshop of Discourse and Dialogue (SIGDIAL '03)*, pp. 1–12, Sapporo, Japan, July 2003.
- [20] M. S. M. Rahim, A. Rehman, S. Ni'matus, F. Kurniawan, and T. Saba, "Region-based features extraction in ear biometrics," *International Journal of Academic Research*, vol. 4, no. 1, pp. 37–42, 2012.
- [21] N. Yokokura and T. Watanabe, "Layout-based approach for extracting constructive elements of bar-charts," in *Graphics Recognition Algorithms and Systems*, K. Tombre and A. Chhabra, Eds., vol. 1389, pp. 163–174, Springer, Berlin, Germany, 1998.

- [22] Y. Zhou and C. L. Tan, "Learning-based scientific chart recognition," in *Proceedings of the 4th IAPR International Workshop on Graphics Recognition (GREC '01)*, pp. 482–492, 2001.
- [23] Y. P. Zhou and C. L. Tan, "Hough technique for bar charts detection and recognition in document images," in *Proceeding of the International Conference on Image Processing (ICIP '00)*, vol. 2, pp. 605–608, Vancouver, Canada, September 2000.
- [24] W. Huang, C. L. Tan, and W. K. Leow, "Model-based chart image recognition," in *Graphics Recognition. Recent Advances and Perspectives*, J. Lladós and Y.-B. Kwon, Eds., vol. 3088, pp. 87–99, Springer, Berlin, Germany, 2004.
- [25] A. Mishchenko and N. Vassilieva, "Model-based chart image classification," in *Advances in Visual Computing*, G. Bebis, R. Boyle, B. Parvin et al., Eds., vol. 6939 of *Lecture Notes in Computer Science*, pp. 476–485, Springer, Berlin, Germany, 2011.
- [26] W. Huang and C. L. Tan, "A system for understanding imaged infographics and its applications," in *Proceedings of the ACM Symposium on Document Engineering*, pp. 9–18, Manitoba, Canada, August 2007.
- [27] M. M. Hassan and W. Al Khatib, "Similarity searching in statistical figures based on extracted meta data," in *Proceedings of the Computer Graphics, Imaging and Visualisation (CGIV '07)*, pp. 329–334, Bangkok, Thailand, August 2007.
- [28] L. Yang, W. Huang, and C. L. Tan, "Semi-automatic ground truth generation for chart image recognition," in *Proceedings of the 7th international conference on Document Analysis Systems*, Nelson, New Zealand, 2006.
- [29] G. D. Tourassi, "Journey toward computer-aided diagnosis: role of image texture analysis," *Radiology*, vol. 213, no. 2, pp. 317–320, 1999.
- [30] D. A. Chandy, J. S. Johnson, and S. E. Selvan, "Texture feature extraction using gray level statistical matrix for content-based mammogram retrieval," *Multimedia Tools and Applications*, vol. 72, no. 2, pp. 2011–2024, 2014.
- [31] F. Han, H. Wang, B. Song et al., "A new 3D texture feature based computer-aided diagnosis approach to differentiate pulmonary nodules," in *Medical Imaging: Computer-Aided Diagnosis*, Proceedings of SPIE, San Diego, Calif, USA, 2014.
- [32] F. Han, H. Wang, B. Song et al., "Efficient 3D texture feature extraction from CT images for computer-aided diagnosis of pulmonary nodules," in *Proceedings of the SPIE, Medical Imaging: Computer-Aided Diagnosis*, vol. 9035, pp. 1–7, 2014.
- [33] S. M. Eissen, B. Stein, and M. Kulig, "Plagiarism detection without reference collections," in *Advances in Data Analysis*, R. Decker and H.-J. Lenz, Eds., pp. 359–366, Springer, Berlin, Germany, 2007.
- [34] J. Kasprzak, M. Brandejs, and M. Křipač, "Finding plagiarism by evaluating document similarities," in *Proceedings of the 3rd Workshop on Uncovering Plagiarism, Authorship and Social Software Misuse (PAN '09)*, pp. 24–28, Donostia, Spain, September 2009.
- [35] C. Basile, D. Benedetto, E. Caglioti, G. Cristadoro, and M. D. Esposti, "A plagiarism detection procedure in three steps: selection, matches and squares," Donostia, Spain, 2009.
- [36] Z. A. Hamed and S. Z. Mohd Hashim, *Hybrid particle swarm optimization and black stork foraging for functional neural fuzzy network learning enhancement [UTM Thesis]*, 2012.