Classification and Retrieving Printed Arabic Document Images Based on Bagged Decision Tree Classifier

Matheel E. Abdulmunim[1]**,PhD (Asst. Prof.)** **Haithem K. Abass**[2]

Abstract: Printed Arabic document image retrieval is a very important and needed system for many applications including electronic archiving, search engines, and document management systems. In this paper, an adaptive header-words based printed Arabic document images classification and retrieval system has been proposed that based on decision tree classifier improved by bagging technique. The proposed system implements effective preprocessing and segmentation techniques to prepare the document and correctly detect a specific Arabic header words form query document. Besides that, a collection of discriminative features has been extracted from detected header words to correctly classify them to a right class. In the proposed system, bagging technique has been adapted with decision tree classifier to enhance the performance of classification and hence improve the precision of retrieving documents. The experimental tests confirmed that the proposed system achieved very satisfied results of 97.35% for precision.

Keywords: DIR, header-words, features extraction, bagging, bagged decision tree.

[1] Computer Science Department, University of Technology, Baghdad, Iraq
[2] Software Engineering and Information Technology Department, Al-Mansour University College, Baghdad, Iraq

1. Introduction

Recently, the need for Document Image Retrieval (DIR) systems are widely increased because of the great spread of electronic devices that facilitate the acquisition and archiving of documents. The modern DIR system aims to find relevant documents from a corpus of digitized pages depending on the image features only [1]. Generally, the user submits to the system a query and the result will obtained as a list of documents that match the query in specific features [2]. Many approaches have been proposed for searching and retrieving of documents collection that based on extracting textual features from document image. Lu et al. [3] propose a DIR system that is capable of retrieving English document images by capturing the document content through annotating English word image by a word shape code. The code represents the shape features of Latin character including character holes, and character reservoirs, character ascenders and character descenders. The performance of this English word based documents retrieval system was evaluated by 252 documents and achieved 94.88% of precision. T. Sari and A. Kefali [4] developed a system for indexing and searching Arabic degraded and historical document images without recognizing the textual patterns. The proposed approach deals with textual features of Arabic documents (handwritten or printed). The global features are extracted from each connected component that represent word or sub-word after detection baseline and median zone in the Arabic text. The retrieval system was tested by 123 historical Arabic documents obtained from various resources and the experimental tests show that the proposed system achieved a precision approximating 77.78%. K. Zagoris et al. [5] proposed a DIR system that search and retrieve printed English documents according to a query word image. The proposed retrieval system was based on extracting six significant features that describe the contour and the region shape of each segmented English word image in a document. The performance of the proposed DIR system was tested by a collection of 100 document images obtained from different sources and the system achieved 87.8% of precision. R. Tavoli et al. [6] developed a feature weighting technique for DIR system based on keyword spotting. For each segmented English word, an important seven shape features are extracted. These features describe the English words in terms of area, projection, and geometrical moments. The performance of the proposed DIR system was tested by a dataset of 100 document images from various digital text documents and the system achieved 92.1% of precision. A. Batura et al. [7] proposed a retrieval system for Uyghur document images that based on extracting

Scale-Invariant Feature Transform (SIFT) from Uyghur printed text of Arabic alphabet. The proposed retrieval system was tested by 1000 of different printed Uyghur document images scanned from various resources and the experimental results indicate that the system was achieved 96.29% of average precision.

2. The Proposed System

In many Arabic documents, such as letters, announcements, forms, reports, and other official documents, header words are the main and important text in the documents that may indicate specifically the origin of these documents. Whereas, the header-words in each Arabic document can refer to specific source of organizations such as ministries, universities, colleges, departments, banks, and other institutes. For these important reasons, header-words based retrieval system has been proposed and implemented to retrieve various type of Arabic document images from a dataset of document images. Figure 1 illustrates the structure of header-words based retrieval system that involve a sequence of different steps and each of which has its role in the retrieval operation.

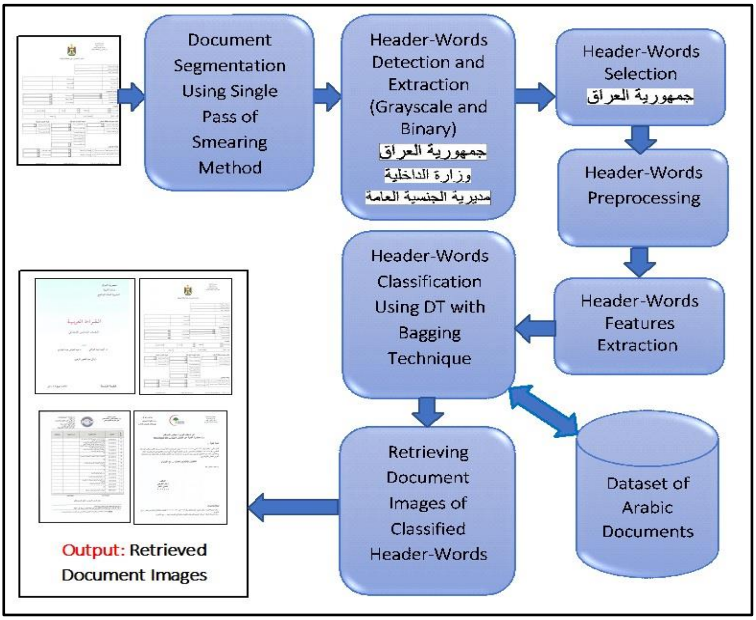
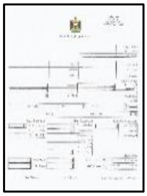


Figure 1: Main structure of the Proposed System.

2.1 Dataset

An official printed Arabic document dataset has been constructed and tested in the proposed retrieval model. This dataset consist of various forms of official printed Arabic document images obtained from different authorized web sites like ministries, universities, government institutions and other official states. The dataset represents many types of Arabic documents like letters, reports, books, forms, announcements, administrative instructions and other official documents. All these papers should have Arabic header words and may contain other objects such as texts, logos, borders, graphics, and other objects. These documents may store in printable format as Portable Document Format (PDF), or Microsoft word document (DOC). The pages of Arabic documents are printed using HP Deskjet printer 2540 series and then scanned by scanner with 300dpi and 600dpi resolutions. In this dataset, 1230 pages are printed and scanned in portrait orientation and in landscape orientation. These scanned document images are stored in three types of color level, first level is black and white image of 1-bit per pixel, second level is grayscale image of 8-bits per pixel, and third level is true color image of 24-bit per pixel. After scan operation, each document image in dataset is stored as a file of JPG (Joint Photographic Group) file format. Table 1 shows samples of Arabic document images dataset of different categories. Table 2 shows the classification of constructed Arabic documents dataset according to their colors and resolutions.

Table 1: Samples of document images in the dataset.

Document Class	Image Sample	Document Class	Image sample
Report		Form	
Book	 القراءة العربية	newspapers	

Letter		Administrative instructions	

Table 2: Samples of document images in the proposed dataset.

Document Class	Image Color and Resolution						Total
	True Color		Grayscale		Black and White		
	300 dpi	600 dpi	300 dpi	600 dpi	300 dpi	600 dpi	
Report	57	8	65	4	80	6	220
Book	18	4	14	4	8	2	50
Letter	160	8	140	4	145	3	460
Form	85	2	85	4	180	2	358
Newspapers	4	1	4	0	3	0	12
Instructions	60	6	36	4	20	4	130
Total	384	29	344	20	436	17	1230

2.2 Segmentation

After converting document image into a binary image using Otsu method [8], word level segmentation has been implemented using modified Run Length Smearing Algorithm (RLSA) [9]. The modifications come from that RLSA technique is applied in horizontal and vertical directions with different variable threshold values and with constant factor that will enhance the smearing operation. These threshold values are computed to control the number of sequence of pixels that will be smeared in the image. For the proposed approach, bounding box is constructed for each connected component in a binary document and then histogram is computed to estimate the value of smeared threshold value. Figure 2 shows the main steps of segmentation algorithm with horizontal and vertical smearing operations for a portion of Arabic document. In horizontal smearing, a number of words with their components are merged together as a black region. In vertical smearing a number of Arabic characters points are combined with their related characters in the words. Making logical

AND between horizontal and vertical smeared objects will eliminate gaps between them.

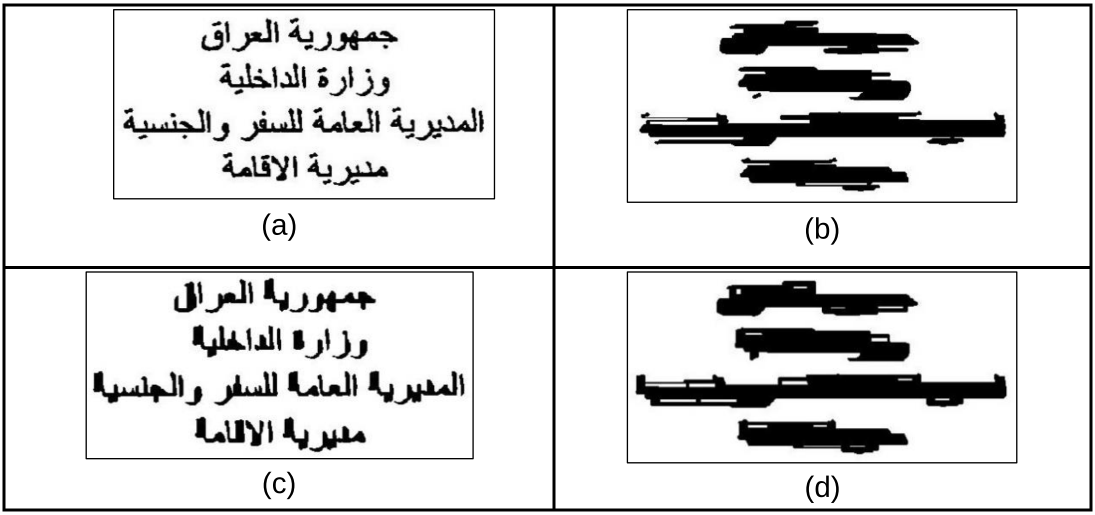


Figure 2: Document segmentation using RLSA technique. (a) Part of binary document image, (b) Image after horizontal smearing, (c) Image after vertical smearing, (d) Image after logical AND between horizontal and vertical smeared images.

2.3 Preprocessing

This preprocessing step is designed specially to enhance the header-words binary image extracted previously in header-words extraction step depending on significant block features to distinguish header words from other objects in document image. It is essentially to reduce the unnecessary superimposed data and keep as much as possible significant information as presented. The preprocessing includes image normalization, outliers removal, and image thinning as shown in figure 3. The preprocessing of header-words will produces thinned header-words image that will be used for the features extraction method in the next stage.

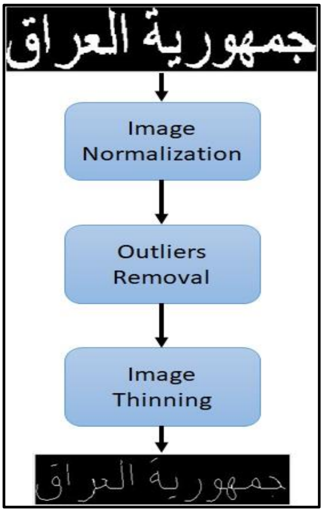


Figure 3: Header-words preprocessing steps.

2.3.1 Image Normalization

The printed Arabic images in the used dataset have various sizes and resolutions. The retrieval systems are sensitive to small variations in the size and position, as in the case of matching templates and correlation methods. Normalization of images size seeks to reduce variations between images due to the size of Arabic words to improve the performance of the classifier. Therefore, in the proposed system all the images are normalized into size 64 of the height and make the width flexible according to the image content to preserve the aspect ratio of the sentence shape. An example of the normalization process is shown in figure 4.

Figure 4: Image normalization. (a) Binary image, (b) Normalized image.

2.3.2 Outliers Removal

During the image acquisition process, some of false pixels are appeared in the image. These pixels called outliers, which affect the results of the feature extraction method as well as affecting the whole system performance. Therefore, a proposed algorithm is used to remove unwanted pixels from the binary image as illustrated in figure 5. The outliers are removed by selected threshold value which represent the smelliest possible component in the Arabic language. Therefore, any component is smaller than the selected threshold is considered as an outlier.

Figure 5: Outliers removal. (a) Binary image, (b) Image after removing outliers.

2.3.3 Image Thinning

Thinning algorithm is a morphological operation that is used to remove selected foreground pixels from binary image. It preserves the topology (extent and connectivity) of the original region while throwing away most of the original foreground pixels. In the header-words preprocessing module, Zhang-Suen thinning algorithm [10] is applied to binary header-words image BHW_{image} and the result of applying this thinning algorithm is thinned header-words image THW_{image} shown in figure 6.

Figure 6: Image thinning. (a) Original image, (b) Thinned image.

2.4 Header-Words Features Extraction

In this stage, appropriate and significant features have been extracted from the header-words images to make essential description for them. A combination of two different types of features are extracted from the header-words. These features are structural features and statistical features.

A. Structural Features

Primitive visual Arabic writing are part of the structural descriptors that are connected in the form of writing. Various important descriptors can be extracted from Arabic scripts as shown in figure 7. In this stage, a number of considerable structural features are extracted from the header-words image resulted from preprocessing stage of the proposed system. The extracted features are number of dots, number of loops, number of end points, number of junctions number of vertical lines.

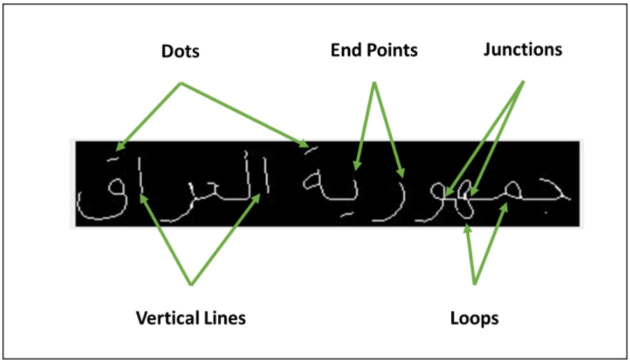


Figure 7: The extracted structural features from Arabic header-words.

1. **Number of dots:** It is very important features, due to the variances in the Arabic text. The dots are the smallest items in the header-words image, and it contains only a little pixels. In order to distinguish between Arabic words that have the same number of dote but may represent different header-words, the number of upper dots (NUD) that are located above the baseline and lower dots (NLD) that are located under the baseline have been computed.

2. **Number of loops:** Another important structural feature that was extracted is the number of loops in the header-words. Nine of the Arabic alphabets has one loop and only one character has two loops. Loops feature has a better effect on describing a text image when it is extracted from an Arabic words than extracted from a single character. Counting the number of loops (NL) in the characters that constitute the header-words will provide a useful characterization of the Arabic words and help to differentiate between them.

3. **Number of end-points:** This significant structural feature can be extracted from Arabic words that have different number of end-points

(NE). In order to detect these points, eight templates are proposed as shown in figure 8. Where, the number '1' in the center of each template will represents the end point in the thinned image.

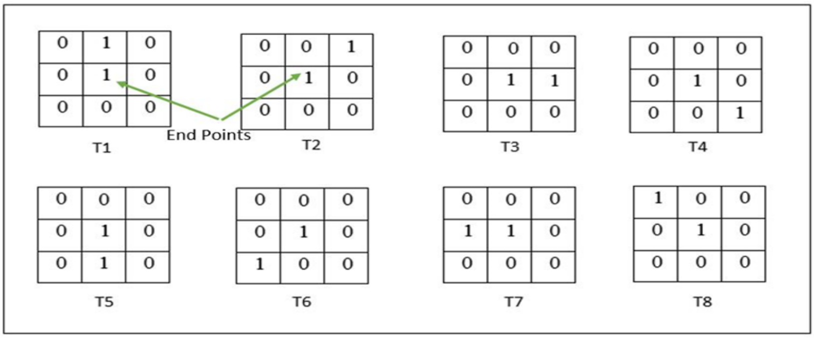


Figure 8: Eight templates to detect the end-points.

4. **Number of junctions:** This feature can also help in describing the global characteristics Arabic words image. In order to detect and extract number of junctions (NJ), four templates are proposed as shown in Figure 9. Where the number '1' in the center of each templates represents the junction point.

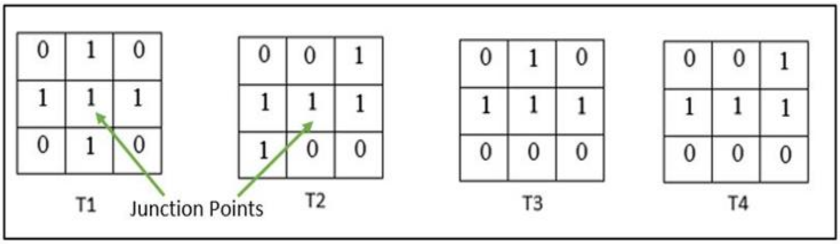


Figure 9: Four templates to detect the junction points.

5. **Number of vertical lines:** An expressive feature that represents the number of vertical lines (NV) is extracted from grayscale header-words image. The proposed feature gives appropriate description for header-words images to make a good distinguish among them. In this feature, the direction and magnitude of image gradient are obtained from the grayscale header-words. The magnitude of header-words image gradient is shown in figure 10.

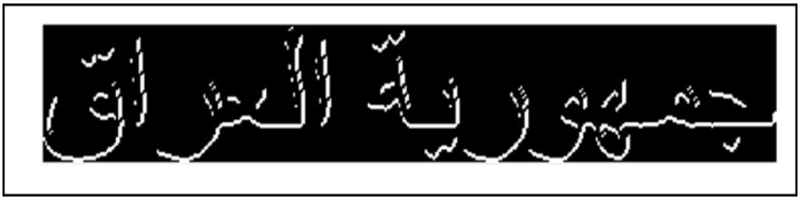


Figure 10: Magnitude of header-words image gradient.

The results of applying direction of image gradient will produce an array within range from -π to π. In the proposed feature, it is necessary to detect only the possible vertical lines in the text image. Since the vertical lines have angle 90^0 and some of lines are curved, the proposed algorithm used the range 60^0 to 120^0 to detect vertical lines in image gradient. The result of selecting these directions is illustrated in figure 11.

Figure 11: Image after selecting the range 60^0 to 120^0 of image direction.

After that, only parts of the text with the chosen directions are appear in the output image. The next step is to keep only the required lines by eliminate the small parts in the gradient image and keep the bigger ones as shown in figure 12.

Figure 12: The detected vertical lines in header-words image.

Finally, these structural features are combined into a single structural features vector $V_{structural}$={NUD, NLD, NL, NE, NJ, NV}.

B. Statistical Features

There are a number of statistical features can be detected specifically for Arabic scripts. In the proposed system of these features are extracted, which are number of components in the sentence, number of components, upper profile points, and left profile points.

1. **Number of components:** The Arabic header-words can be considered as a text image that consists of a number of connected components (NC). These components can represent characters, sub words, words, and dots as shown in figure 13. The number of these components will be extracted from the thinned header words image as a statistical feature.

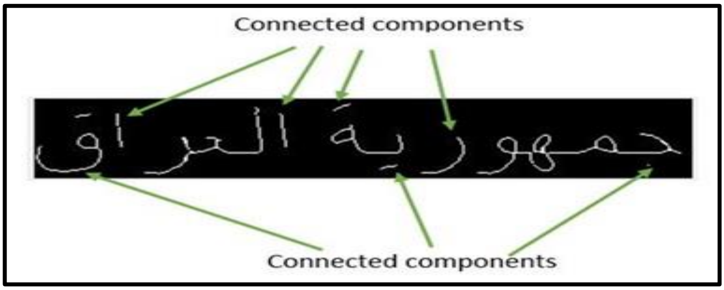


Figure 13: Connected components in the header-words image.

2. Upper and left profile points: An appropriate statistical feature has been proposed that based on the upper profile and the left profile points of the Arabic words image as illustrated in figure 14. The upper profile points are located firstly by computing the number of background pixels from upper border (N_UP) to nearest foreground pixel in the header-words image. In addition, the left profile points are located by computing the number of background pixels from left border (N_LEFT) to nearest foreground pixel in the text image.

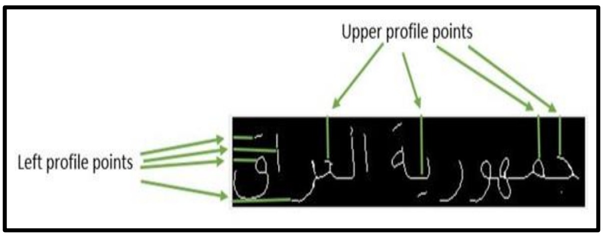


Figure 14: Upper profile and left profile points in header-words image.

Finally, these statistical features are combined into a statistical features vector $V_{statistical}$={NC, N_UP, N_LEFT}.

2.5 Classification Using Decision Tree with Bagging Technique

Classification is analytical modeling and it is a method of supervised learning in which the class target is already known. The classification techniques has two parts: the first part is learning practice, where training data will be analyzed. The learned type is characterizing in the way of classification regulations. The second part of classification is testing practice, in which test information to calculate approximately the exactness of classifier [8].

Decision Tree (DT) is a predictive model, which converts observations about an object to conclude about the object's target value. DT algorithm is a classification technique that recursively partitions a dataset of records using depth-first greedy approach until all the data items belong to a particular class. Each path in decision tree forms a decision rule that utilizes greedy approach from top to bottom. DT has many advantages as summarized below:

- Simplicity and fast classification.
- Produces results with high accuracy.
- Representation is easy to understand.
- It supports incremental learning.
- Less memory requirements.

However, DT technique suffer from high computation time and over fitting problem [11].

Bootstrap Aggregating (Bagging) is an algorithm that was designed to improve the stability and accuracy of machine learning algorithms. It can be used with many regression and classification methods to reduce the variance associated with prediction, and this will improve the prediction process [12]. Bagging is a voting technique however base-learners are made different by training them over slightly different training sets. The basic idea is that many bootstrap samples are depicted from the available data, some prediction algorithm is applied to each bootstrap sample, and then the results are collective, by simple voting for classification, to gain the overall prediction, with reducing variance due to the averaging. Bagging is the application of the bootstrap procedure to a high-variance machine learning algorithm, typically decision trees. Bagging has a single parameter, which is the number of trees. All trees are unpruned binary tree and at each node in the tree one searches over all features to find the feature that best splits the data at that node [13]. The overall process of bagging is shown in figure 15.

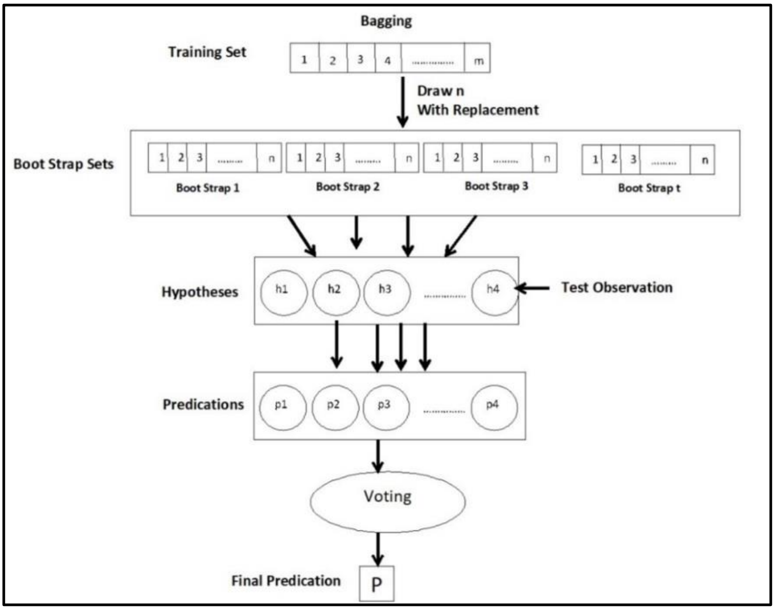


Figure 15: Main steps of bagging technique.

In the proposed retrieval system, classification technique plays significant role in identifying to which of a set categories a new header-word image belongs, based on a training set of data containing instances whose category membership is known. Each header-words image is represented as a features vector that combine the statistical and structural into a single vector to make a clarify description about each header-words image. The combined feature vector ($V_{combined}$) is one-dimensional array that consist of 9 features (6 for structural and 3 for statistical). In the proposed system, DT classifier improved by bagging technique (Bagged DT) has been proposed. The classification process involves two phases, training phase and testing phase.

1. **Training phase:** In this phase, each class of Arabic header-words image (represented by $V_{combined}$) is given a unique label to represent it and distinguish it from the other classes of header-words image. Then, DT classifier will be trained with the aid of bagging technique

by assigning all the images in the training set that have the same header-words into its desired label. Therefore, the output of the training process are training models that will be used later in testing phase. The process of training is illustrated in algorithm 1.

Algorithm 1: ***Training Phase.***
Input: Training set (set of combined feature vectors $V_{combined}$). **Output**: Training models.
Begin **Step 1:** Load the training set. **Step 2:** Assign unique class label for the each feature vector $V_{combined}$ in the training set that represent the same header-words image. **Step 3**: Collect all the feature vectors of the same class label as a group. **Step 4:** Create a number of subsets of training data randomly from all groups, where the created subsets should cover all training data. **Step 5:** For each class label do the following: **5.1:** Train each created subset within DT classifier. **5.2:** Save the classifier result as a model. **5.3:** Combine the resulted model into a set training model. **Step 6**: Return set of training models. **End.**

2. **Testing phase:** In this phase, the feature vector $V_{combined}$ that represent a header-words image will be tested by passing through all the training models. In this case, each training model will classify the testing feature vector by assigning it to a specific label. After each training model gives the testing vector a predicted label, then voting process of bagging technique is used in order to choose the most frequent predicted results that appeared from the training models and make it as the final classification decision. The process of testing is illustrated in algorithm 2.

Algorithm 2: ***Testing Phase.***
Input: Testing vector (combined feature vector $V_{combined}$). **Output**: Predicted class label.
Begin **Step 1:** Load the testing vector ($V_{combined}$) that represent a header-words image. **Step 2:** Load the training models resulted from training phase (algorithm 2.24). **Step 3:** For each training model do the following:

3.1: Pass testing vector ($V_{combined}$) to the current training model to classify it to a certain class label.
3.2: Save the result of classifier.
Step 4: Apply the voting technique to the classifier results from all training models.
Step 5: Save the classifier result with high voting as a predicted class label.
Step 6: Return the predicted class label.
End.

2.6 Retrieving Document Images

The last step of in the header-words based retrieval system is retrieving the desired Arabic documents from the dataset system according to the query header-words input. The output of bagged DT algorithm is a class label that represent one of the Arabic header-words image. Thus, all the documents which have the same header-words input are retrieved from the dataset of Arabic document images and displayed.

3. Tests and Experimental Results

The proposed system was implemented in Matlab 2016a and Visual Studio 2013 programming languages. The experiments were performed on an Intel Core i7, 64 bit Operating System, 2.50 GHz processor and 6GB RAM. However, in the proposed system bagged DT was proposed to achieve more accurate classification than traditional DT and hence improved the performance of document retrieving. Table 3 shows performance comparison between the retrieving system based on DT classifier and the proposed system that based on bagged decision tree classifier. The performance evaluation depends on the following evaluation metrics [14]:

1. **Precision:** Represents the percentage ratio of retrieved document images relevant to a given query to the total number of retrieved documents as defined in equation 1.

$$precision = \frac{Number\ of\ relevent\ documents\ retrieved}{Number\ of\ retrieved\ documents} \times 100 \qquad (1)$$

2. **Recall:** Represents the percentage ratio of retrieved document images relevant to a given query to the total number of relevant document images in dataset as defined in equation 2.

$$recall = \frac{Number\ of\ \ relevent\ documents\ retrieved}{Number\ of\ relevemt\ documents\ in\ dataset} \times 100 \qquad (2)$$

3. **F-score:** Represents the harmonic mean between recall and precision values as defined in equation 3.

$$f - score = \frac{2 \times precision \times recall}{precision + recall} \qquad (3)$$

Table 3: Performance comparison between DT and bagged DT classification techniques.

Classification Technique	Precision	Recall	F-score
DT	91.58%	90.05%	90.81%
Bagged DT (proposed)	97.35%	96.93%	97.13%

The table of comparison verified that the proposed bagged DT technique was noticeably enhanced the precision of retrieving documents. The enhancement was due to the accurate classification of bagged DT because of its ability to build a number of predicted models based on bootstrap aggregation in tanning phase and voting technique which helped for predicting of right class in testing phase. This enhancement has improved the stability and reduced the variance of classification.

4. Conclusions

An efficient printed Arabic document classification and retrieval system was proposed in this paper. The proposed system retrieves the desired documents based on extracting appropriate features from header-words of Arabic document. The extracted features is a combination of structural and statistical which are correctly discriminate each header-word from the others. Furthermore, an accurate classification method based on DT improved by bagging technique was implemented and achieved a very satisfied results. In addition, a real and comprehensive printed Arabic document database within 70% for training and 30% for testing was constructed to evaluate the performance of proposed document retrieval system. The table of results confirms that the proposed bagged DT was considerably enhanced the precision of retrieving documents. The enhancement was due to the more accurate classification of bagged DT than traditional DT.

References

[1] F. Alaei, A. Alaei, M. Blumenstein, and U. Pal, "A Brief Review of Document Image Retrieval Methods: Recent Advances," International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 3500-3507, 2016.F.

[2] M. K. I. Kassab, S. S. A. Naser, and M. J. A. Shobaki, "An Analytical Study of the Reality of Electronic Documents and Electronic Archiving in the Management of Electronic Documents in the Palestinian Pension Agency (PPA)," European Academic Research, vol. 4, no. 12, pp. 10052-10102, 2017.

[3] S. Lu, a. Linlin Li, and C. L. Tan, "Document Image Retrieval Through Word Shape Coding," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 11, pp. 1913-1918, 2008.

[4] T. Sari, and A. Kefali, "A Search Engine for Arabic Documents," Colloque International Francophone sur l'Ecrit et le Document, pp. 97-102, 2008.

[5] K. Zagoris, K. Ergina, and N. Papamarkosa, "A Document Image Retrieval System," Engineering Applications of Artificial Intelligence, vol. 23, no. 6, pp. 872-879, 2010.

[6] E. K. Reza Tavoli, M. Shojafar, H. Soleimani, and Z. Pooranian, "Weighted PCA for Improving Document Image Retrieval System Based On Keyword Spotting Accuracy," 36th International Conference on Telecommunications and Signal Processing (TSP), IEEE, pp. 773-777, 2013.

[7] A. Batura, G. Tursuna, M. Mamutb, N. Yadikara, and K. Ubula, "Uyghur Printed Document Image Retrieval Based on SIFT Features," International Congress of Information and Communication Technology (ICICT 2017), pp. 737-742, 2017.

[8] C. Nabendu, S. H. Shaikh, and K. Saeed, "A Comprehensive Survey on Image Binarization Techniques," Exploring Image Binarization Techniques, Springer, pp. 5-15, 2014.

[9] M. E.Abdulmunim, and H. K. Abass, "Logo Detection in Arabic Documents Using Multi Smearing Method and Decision Tree," Al-Mansour Journal, Issue 27, pp. 1-14, 2017.

[10] W. Chen, L. Sui, Z. Xu, and Y. Lang, "Improved Zhang-Suen Thinning Algorithm in Binary Line Drawing Applications," *International Conference on Systems and Informatics (ICSAI),* IEEE, pp. 1947-1950, 2012.

[11] M. Soundarya, and R. Balakrishnan, "Survey on Classification Techniques in Data mining," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 3, no. 7, pp. 7550-7552, 2014.

[12] E. Alpaydın, "Introduction to Machine Learning," Second Edition, MIT Press, 2010.

[13] F. Barboza, H. Kimura, and E. Altman, "Machine Learning Models and Bankruptcy Prediction," *Expert Systems with Applications*, vol. 83, pp. 405-417, 2017.

[14] U. D. Dixit, and M. S. Shirdhonkar, "A Survey on Document Image Analysis and Retrieval System," *International Journal on Cybernetics & Informatics (IJCI),* vol. 4, no. 2, pp. 259-270, 2015.

تصنيف واسترجاع الوثائق العربية المطبوعة بالاعتماد على المصنف شجرة القرار *Bagged*

ا. م. د. مثيل عماد الدين عبدالمنعم هيثم كريم عباس

المستخلص: إن استرجاع صورة الوثائق العربية المطبوعة هو نظام مهم جدا ومطلوب للعديد من التطبيقات بما في ذلك الأرشفة الإلكترونية ومحركات البحث وأنظمة إدارة الوثائق. في هذه البحث تم اقتراح نظام متكيف لتصنيف واسترجاع الوثائق العربية المطبوعة استنادا الى الكلمات الرأسية بالاعتماد على طريقة شجرة القرار(Decision Tree) التي تم تحسينها بواسطة تقنية Bagging. يقوم النظام المقترح بتطبيق تقنيات المعالجة المسبقة والتجزئة الفعالة لإعداد الوثيقة وكشف الكلمات الرأسية من وثيقة الاستعلام وبالإضافة إلى ذلك، تم استخراج مجموعة من الميزات التمييزية من الكلمات الرأس المستخرجة لتصنيفها بشكل صحيح إلى الصنف الصحيح. في النظام المقترح، تم تكييف تقنية Bagging مع المصنف شجرة القرار لتعزيز أداء التصنيف وبالتالي تحسين دقة استرجاع الوثائق. اثبتت التجارب ان النظام المقترح حقق نتائج مقنعة جدا في الدقة وكانت 97.35%.

الكلمات المفتاحية: رأس الكلمات، ميزات الاستخراج، التعبئة، شجرة القرار المعبأة.

[1] قسم علوم الحاسب، الجامعة التكنولوجية، بغداد، العراق
[2] قسم هندسة البرمجيات وتكنولوجيا المعلومات، كلية المنصور الجامعة، بغداد، العراق