Arabic Artistic Script Style Identification Using Texture Descriptors.

Zineb Kaoudja
LAGE Laboratory
university of Kasdi Merbah
Ouargla, Algeria
zinebkaoudja@gmail.com

Belal Khaldi
LAGE Laboratory
university of Kasdi Merbah
Ouargla, Algeria
bilel.khaldi@gmail.com

Mohammed Lamine Kherfi
LAMIA laboratory
université du Québec
à Trois-Rivières, Canada and
LAGE laboratory
university Kasdi Merbah
Ouargla, Algeria
Kherfi@uqtr.ca

***Abstract*— Texture descriptors have been widely used for many computer vision tasks. Document Analysis (DA) is a hot topic of research, which aims at analyzing digitized documents containing different types of texts. In this paper, we design and evaluate a system for Arabic artistic style recognition using numerous texture descriptors. These descriptors have been chosen based on their effectiveness and the high performance they showed in other related works. To make it more comprehensive, the system has been evaluated using five different classifiers. Results indicated that texture descriptors appropriately fit the task and yielded almost perfect performances for most of the styles.**

***Keywords*— Arabic calligraphy handwriting; texture descriptors; feature extraction; Arabic Optical Font Recognition (AOFR).**

I. INTRODUCTION

Digital Image Analysis (DIS) is the set of techniques that are used to represent the content of images in an appropriate manner. This representation can be used thereafter for recognizing, differentiating, and quantifying diverse types of images. DIS techniques have widely been used in several application areas such as robotics, handwriting recognition, document analysis, image indexing and retrieval, diagnostic assistance in medicine, object detection, pattern recognition, etc.

Document Analysis (DA) is a hot topic of research, which aims at analyzing digitized documents containing different types of texts such as numbers [1], letters [2, 3], words [4], sentences or paragraphs [5]. Frequently, DA systems utilize both DIS and machine learning techniques in order to automatize their tasks. As illustrated in Fig. 1, DA systems involve a set of major including pre-processing, feature extraction, and classification. The purpose of each of these steps can be resumed as follow:

- Pre-processing: aims to de-noise or crop the image to remove useless information; threshold/binarize the image to highlight objects or separate foreground from the background.

- Feature extraction: is about extracting the most significant information from the image. The result of this step is a value, vector or matrix that can be used later on to discriminate images.

- Classification: This step consists of two phases, a) training: in which we teach our model the common properties of each class of images by jointly providing the features and labels of each class. b) Classifying: this phase consists of feeding unlabeled images to the trained model in order to assign them to their appropriate classes (i.e., labeling them).

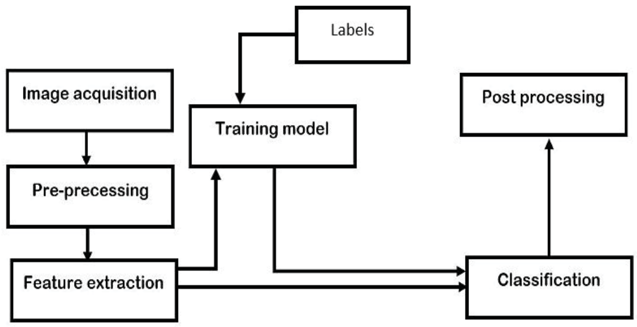


Fig. 1. A common architecture of digital document analysis (DA) systems.

Feature extraction is a key step that is substantially addressed in many image classification systems including DA. Features can be defined as the individual measurable heuristic properties of the phenomena being observed [6]. A considerable number of image features have been proposed in literature based on different approaches such as color, texture, shape, points, edges or object-based features. The choice of the features to be extracted is guided by the image content and the needed type of information.

As for digitized DA systems, images contain a text composed of a finite number of letters (e.g., 28 letters in Arabic) and sometimes diacritics as in the Arabic language. Thus, a digitized document image will comprise repetitions of the same or similar letters and diacritics. Texture features have therefore widely been used in DA systems in which the image can be decomposed in patterns, either repeating or non-repeating [7].

In a former work [24], we have proven that texture is a powerful feature that may outperform even deep-learning-based features in several datasets. Numerous solutions for DA systems have been established based on texture descriptors. In [8], we proposed a new system for Arabic calligraphy style recognition based on Local Phase Quantization (LPQ) and a combination of three different classifiers namely K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). F. Alaie & al. performed a comparative study

978-1-7281-5835-8/20/$31.00 ©2020 IEEE

between twenty-six texture descriptors to demonstrate their performance on three public datasets [9]. A proposed solution from S.Ghosh & al [10] utilized LBP descriptor to isolate non-text from text components in handwritten documents. In [11], authors have proposed a technique for writer identification using Histogram of Oriented Gradients (HOG) and Gray Level Run Length (GLRL) Matrices.

As we have mentioned, texture features have extensively been used, in literature, for DA systems. However, there is no former work that addresses the performance of texture descriptors in Arabic calligraphy/artistic documents classification. This work aims to evaluate some of the most powerful texture descriptors for DA purposes. These descriptors are Gray Level Co-occurrence Matrix (GLCM) [12], Local Binary Pattern (LBP) [13], Weber Local Descriptor (WLD) [14], Histogram of Oriented Gradient (HOG) [15], Local Phase Quantization (LPQ) [16] and Binarized Statistical Image Features (BSIF) [17]. The choice of these descriptors has been performed based on the effectiveness they showed in different other application areas such as biometry [18, 19, 20] for LBP and BSIF, object detection for HOG, handwriting recognition [25] for LPQ, computer-aided diagnostics[26] for WLD and image retrieval[27] for GLCM. However, it should be mentioned that none of the aforementioned descriptors has ever been considered for Arabic calligraphy document classification. For that, our goal is to test the performance of these powerful texture descriptors on Arabic Calligraphy style recognition.

The rest of the paper is organized as follows: In section 2 we introduce the involved texture descriptors. Section 3 holds the experiment results and discussions. Finally, we draw some conclusions.

II. TEXTURE DESCRIPTORS

In this work, we study the influence of using texture descriptors in DA systems for Arabic calligraphy document classification. The focus has been on six of the most powerful and commonly used texture descriptors, which are GLCM, LBP, WLD, HOG, LPQ, and BSIF.

A. GLCM [12]

amongst the recognized and widely used techniques for texture representation. It contains a set of second-order statistics that measures the spatial dependency among gray-levels. Given a displacement vector (Δx,Δy) and an image I with a size of $M \times N$, GLCM can be extracted using Eq. 1.

$$M(p,q) = \sum_{i=1}^{N}\sum_{j=1}^{M}\begin{cases} 1 & if\ I(i,j) = p \\ & I(i+\Delta x, j+\Delta y) = q \\ 0 & otherwise. \end{cases} \quad (1)$$

Haralick [12] has introduced a set of statistical measurements that can be extracted from GLCM. In this work we extract only four statistical measurements of Haralick namely contrast, correlation, energy, and homogeneity, these features are the least correlated together and these four sufficed to give good results in classification.

B. LBP [13]

It is a simple yet very efficient texture descriptor that consists of mapping each pixel of the image, based on its neighborhood, into its corresponding binary code as shown by Eq. 2. The value of the LBP code of a pixel (xc, yc) is calculated using the following formulas.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s\left(g_p - g_c\right)2^p \quad s(x)\begin{cases} 1, if\ x \geq 0; \\ 0, otherwise \end{cases} \quad (2)$$

where P is the number of neighbors points on a circle of radius of R around the central pixel.

The appearance frequency of each binary code is thereafter accumulated into a histogram and considered as the descriptor of the image.

C. WLD [14]

extracts at each pixel level in the image two measures namely, differential excitation and orientation. Differential excitation is the function of the ratio between the intensity of this pixel against its neighbors, whereas, the orientation is the gradient orientation of the corresponding pixel. These two components are respectively given by the following formulas:

$$\xi = \arctan\left(\frac{\Delta I}{I}\right) \quad (3)$$

$$\varphi_t = f_q(\theta') = \frac{2t}{\pi}\pi, t = mod\left(\left[\frac{\theta'}{2\pi/T} + \frac{1}{2}\right]\right) \quad (4)$$

wherein ΔI is the difference in intensity, I is the intensity of the current pixel and θ is the orientation angle at each pixel level.

The appearance frequency of one component given the other is then accumulated in a histogram and consider as the descriptor.

D. HOG [15]

this feature descriptor consists of calculating the appearance frequency of gradient orientation. To extract HOG from a patch image (i.e., 64×128 pixels) the gradient orientation and magnitude are calculated for each pixel p(x,y) using the following formulas respectively:

$$\varphi(x,y) = \arctan\left(\frac{f_y(x,y)}{f_x(x,y)}\right) \quad (5)$$

$$e(x,y) = \sqrt{f_x^2(x,y) + f_y^2(x,y)} \quad (6)$$

where fx(x, y) and fy(x, y) are the differences of brightness in the horizontal and vertical direction respectively.

Subsequently, HOG histogram (9×1) is extracted from each cell, counting the number of occurrences of each orientation, and then concatenated to form the final HOG histogram.

E. LPQ [16]

It is a blur insensitive texture descriptor, essentially invented to deal with image blurring. LPQ method founded on the blur in-variance property of the Fourier phase spectrum [22]. At each pixel position x of the image, the local phase information is computed over M-by-M neighborhoods using a short-term Fourier transform (STFT) using the following equation:

$$F(u,x) = \sum_{y \in N_x} f(x-y) e^{-j2\pi u^T y} = w_u^T f_x \tag{7}$$

Where wu is the basis vector of the 2-D DFT at the frequency u, and f is another vector containing all M2 image samples from Nx.

F. BSIF [17]

computes a binary code for each pixel by linearly projecting local image patches onto a subspace, whose basis vectors are learned from natural images via independent component analysis (ICA) and principal component analysis (PCA), and by binarizing the coordinates in this basis via thresholding. The number of basis vectors determines the length of the binary code string.

$$s_i = \sum_{u,v} W(u,v)\, X(u,v) = w_i^T x \tag{8}$$

where W is the learned filters X is the image window.

III. RESULTS AND DISCUSSION

The aim of this section is to evaluate the performance of GLCM, LBP, WLD, HOG, BSIF, and LPQ for calligraphy image classification. To this end, we use five different classifiers namely Random Forest (RF), Support Vectors Machines (SVM), Decision Tree (DT), K-Nearest Neighbor (KNN), and Local Discriminant Analysis (LDA). The following configurations, which found to be the best, have been set:

A. Classifiers

- For SVM, we choose ‘Polynomial’ kernel, and ‘onevsall’ parameters.
- For KNN, we set $k \in \{1,2,..10\}$.
- For RF, we consider using 500 trees.

B. Features

- GLCM: we use three distances {1, 3, 5 } along four different orientations {0º, 45º, 90º, 135º }. Four statistical moments have been extracted from GLCM namely: contrast, correlation, energy, and homogeneity.
- LPQ: we use “Gaussian derivative quadrature filter pair”, and a window size of 3×3.
- LBP: we use the uniform-LBP, with a 3×3 as the window size.
- WLD: we set the block and the window sizes to 5×5 and 3×3, respectively.
- BSIF: we use 8 texture filters with the size 3×3.
- HOG: we set a 3×3 the number of HOG windows, and 19 the number of the histogram bins.

C. Dataset

Currently, there is no universal Arabic handwriting calligraphy-text dataset for font recognition. Therefore, we use our proposed dataset [8] which comp-rises images of handwritten Arabic calligraphy texts, under different variations such as scale and text length. It consists of 1685 images categorized into 9 different styles with three of these styles, named Mohakik, Maghribi, and Squar-Kufi are used for the first time. Fig. 2. Illustrates representative samples from our Arabic calligraphy dataset.

Fig. 2. Representative images of Arabic calligraphy handwritten texts from our created dataset.

D. Metrics

In our evaluation, we have opted for 3-folds cross-validation. i.o.w. The dataset is divided into 3 sub-sets, wherein each evaluation round one set is used for testing and the two others for training. As for performance metrics, we calculate precision, recall, F1 score, and accuracy:

$$Precision = \frac{true\ positive}{true\ positive + false\ positive} \tag{9}$$

$$Recall = \frac{true\ positive}{true\ positive + false\,negative} \tag{10}$$

$$F1 = 2 * \frac{precision * recall}{precision + recall} \tag{11}$$

$$Accuracy = \frac{true\,positive + true\ negative}{true\,positive + false\,positive + true\ negative + false\ negative}$$

(12)

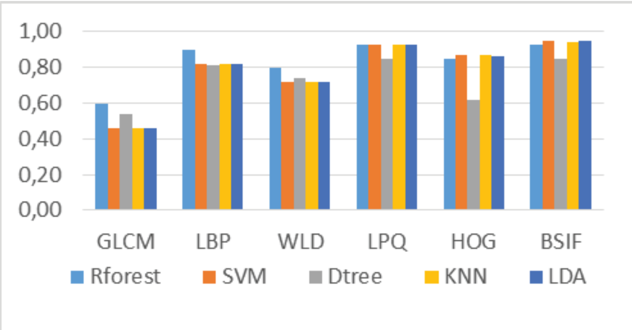
TABLE I. illustrates the obtained mean accuracies yielded by each descriptor using different classifiers. From these results, it is clear that the best combinations descriptor/classifier that provides the highest accuracies are BSIF/SVM with 94.8%, LPQ/SVM with 93.8%, LBP/RF with 89.8%, HOG/SVM with 87.3 %, WLD/RF with 81.1% and GLCM/RF with 60.4%, respectively.

Based on these results, we can say that the best descriptor for Arabic artistic writing recognition is BSIF descriptor.

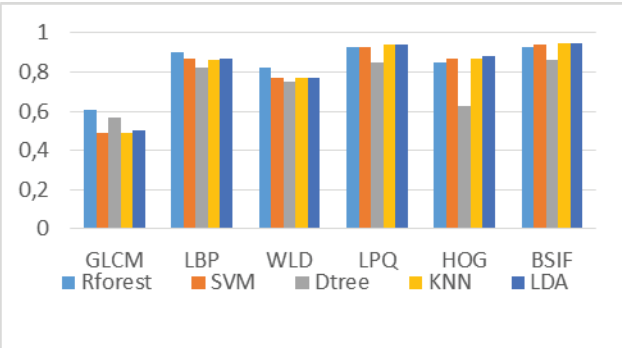
TABLE I. COMPARISON OF THE MEAN ACCURACY RESULTS OF ALL STUDIED TEXTURE DESCRIPTORS WITH FIVE DIFFERENT CLASSIFIERS.

	RF	SVM	DT	KNN	LDA
GLCM	**60.4%**	52.5%	55.2%	57.9%	55.6%
LBP	**89.8%**	84.6%	82%	84.2%	76.4%
WLD	**81.1%**	74.8%	75.4%	69.6%	61.3%
LPQ	93.1%	**93.8%**	85.5%	87.3%	73.8%
HOG	84.9%	**87.3%**	63%	78.3%	85.1
BSIF	93.9%	**94.7%**	85.2%	84.4%	85.9%

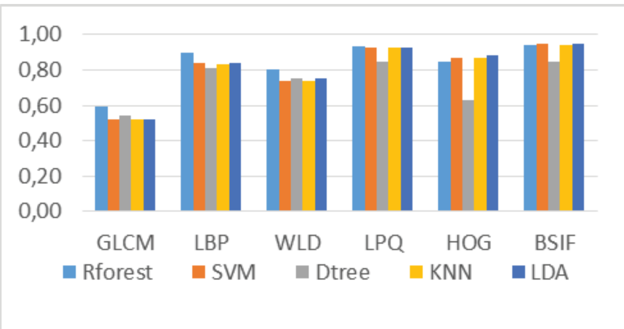
To further confirm these results, we have individually plotted the F1 score measure, precision and recall are shown individually for all descriptors with all the classifiers (Fig. 3. a, Fig. 3. b and Fig. 3. c).



(a)



(b)



(c)

Fig. 3. The performance of all possible descriptor/classifier composites using different metrics, a) F1-Score, b) Precision and c) recall.

From Fig. 3, we can confirm that the best performance has been shown by BSIF followed by LPQ. However, we can also notice that LBP has yielded relatively high results compared to the remaining descriptors. This could be attributed to the capability of these descriptors of recognizing micro-patterns within local neighborhoods and not just rely on the difference in intensity.

It should be mentioned that some artistic styles are hard to recognize than others. This is because some styles share common features such as Thuluth to Mohakik. Therefore, we separately list the performance yielded for each style, using the best descriptor/classifier combination, in TABLE II.

TABLE II. THE YIELDED PERFORMANCE FOR EACH ARTISTIC STYLE SEPARATELY USING BSIF DESCRIPTOR AND SVM CLASSIFIER.

	Precision	Recall	F1-score
Diwani	0.9948	1.0000	0.9974
Naskh	0.9845	1.0000	0.9922
Farisi	0.9714	0.9278	0.9490
Rekaa	0.9371	0.9727	0.9545
Thuluth	0.8919	0.8410	0.8624
Kufi	0.9677	0.9389	0.9503
Maghribi	0.9498	1.0000	0.9728
Mohakik	0.8486	0.8629	0.8525
Square-kufi	1.0000	0.9841	0.9920

From TABLE II., we can see that BSIF/SVM combination has correctly classified all the images belonging to the styles Diwani, Naskh, and Maghribi in terms of Recall. In terms of precision, it seems that Square-kufi, Diwani and Naskh styles have obtained the best results owing to their unique features. To find out which of the styles are misclassified with others, we have calculated and presented the confusion matrix in TABLE III.

TABLE III. CONFUSION MATRIX USING BSIF DESCRIPTOR AND SVM CLASSIFIER.

	1	2	3	4	5	6	7	8	9
1	**1,00**	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2	0,00	**1,00**	0,00	0,00	0,00	0,00	0,00	0,00	0,00
3	0,00	**0,01**	**0,93**	**0,07**	0,00	0,00	0,00	0,00	0,00
4	0,00	0,00	**0,03**	**0,97**	0,00	0,00	0,00	0,00	0,00
5	**0,01**	**0,01**	0,00	0,00	**0,84**	0,00	0,00	**0,15**	0,00
6	0,00	0,00	0,00	0,00	0,00	**0,94**	**0,06**	0,00	0,00

7	0,00	0,00	0,00	0,00	0,00	0,00	**1,00**	0,00	0,00
8	0,00	**0,01**	0,00	0,00	**0,10**	**0,03**	0,00	**0,86**	0,00
9	0,00	0,00	0,00	0,00	**0,01**	0,00	0,00	**0,01**	**0,98**

From this confusion matrix, it is obvious that the highest error rate is 15% (resp. 10%) related to the misclassified images from Thuluth to Mohakik (resp. vice-versa). This occurs because these two styles share many common characteristics that make it hard to distinguish them even for the human eye. As illustrated in Fig. 6, Thuluth and Mohakik have, in most of the cases, the same characteristics, such as the writing style and the diacritics. However, Thuluth artists may exaggerate the letter sizes and draw them in a more overlapping manner as represented in Fig. 7.

Fig. 6. The similarity between (a) Mohakik (b) Thuluth.

Fig. 7. Thuluth style with extra overlapped letters.

Conclusion

The present study addresses for the first time the effect of using texture descriptors for Arabic artistic writing recognition. Six of the well-known and most used texture descriptors have been used along with five different classifiers. The system has been evaluated on different aspects using different metrics. Results indicated that filter-based (pattern-based) descriptors are more effective for such a task. The best performance has been yielded by the BSIF descriptor with the SVM classifier. Furthermore, a detailed analysis has been conducted to find out which of the Arabic styles has the lowest performance. Results indicated that Thuluth to Mohakik styles has obtained the worst performance among the others. This is attributed to the common characteristics shared by these two styles.

References

[1] Zhao, Hui-huang, and Han Liu. "Multiple classifiers fusion and CNN feature extraction for handwritten digits recognition." Granular Computing (2019): 1-8.

[2] Zhang, Xu-Yao, Yoshua Bengio, and Cheng-Lin Liu. "Online and offline handwritten chinese character recognition: A comprehensive study and new benchmark." Pattern Recognition 61 (2017): 348-360.

[3] Lawgali, Ahmed. "A survey on arabic character recognition." International Journal of Signal Processing, Image Processing and Pattern Recognition 8.2 (2015): 401-426.

[4] Al-Helali, Baligh M., and Sabri A. Mahmoud. "Arabic online handwriting recognition (AOHR): a survey." ACM Computing Surveys (CSUR) 50.3 (2017): 33.

[5] Hashemi, Mahmoud Reza, Omid Fatemi, and Reza Safavi. "Persian cursive script recognition." Proceedings of 3rd International Conference on Document Analysis and Recognition. Vol. 2. IEEE, 1995.

[6] Mangai, Utthara Gosa, et al. "A survey of decision fusion and feature fusion strategies for pattern classification." IETE Technical review 27.4 (2010): 293-307.

[7] Efros, Alexei A., and Thomas K. Leung. "Texture synthesis by non-parametric sampling." Proceedings of the seventh IEEE international conference on computer vision. Vol. 2. IEEE, 1999.

[8] Kaoudja, Zineb, Mohammed Lamine Kherfi, and Belal Khaldi. "An efficient multiple-classifier system for Arabic calligraphy style recognition." 2019 International Conference on Networking and Advanced Systems (ICNAS). IEEE, 2019.

[9] Alaei, Fahimeh, et al. "A comparative study of different texture features for document image retrieval." Expert Systems with Applications 121 (2019): 97-114.

[10] Ghosh, Sourav, et al. "Text/non-text separation from handwritten document images using LBP based features: An empirical study." Journal of Imaging 4.4 (2018): 57.

[11] Hannad, Yaâcoub, et al. "Improving Arabic writer identification using score-level fusion of textural descriptors." IET Biometrics 8.3 (2019): 221-229.

[12] Haralick, Robert M., and Karthikeyan Shanmugam. "Textural features for image classification." IEEE Transactions on systems, man, and cybernetics 6 (1973): 610-621.

[13] Ojala, Timo, Matti Pietikäinen, and Topi Mäenpää. "Gray scale and rotation invariant texture classification with local binary patterns." European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2000.

[14] Chen, Jie, et al. "WLD: A robust local image descriptor." IEEE transactions on pattern analysis and machine intelligence 32.9 (2009): 1705-1720.

[15] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." 2005.

[16] Ojansivu, Ville, and Janne Heikkilä. "Blur insensitive texture classification using local phase quantization." International conference on image and signal processing. Springer, Berlin, Heidelberg, 2008.

[17] Kannala, Juho, and Esa Rahtu. "Bsif: Binarized statistical image features." Proceedings of the 21st international conference on pattern recognition (ICPR2012). IEEE, 2012.

[18] Sharma, Ram Prakash, and Somnath Dey. "Quality Analysis of Fingerprint Images Using Local Phase Quantization." International Conference on Computer Analysis of Images and Patterns. Springer, Cham, 2019.

[19] Arashloo, Shervin Rahimzadeh, Josef Kittler, and William Christmas. "Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features." IEEE Transactions on Information Forensics and Security 10.11 (2015): 2396-2407.

[20] Scherhag, Ulrich, et al. "Detecting morphed face images using facial landmarks." International Conference on Image and Signal Processing. Springer, Cham, 2018.

[21] Liu, Li, et al. "Local binary features for texture classification: Taxonomy and experimental study." Pattern Recognition 62 (2017): 135-160.

[22] Liu, Li, et al. "Extended local binary patterns for face recognition." Information Sciences 358 (2016): 56-72.

[23] Ragb, Hussin K., Vijayan K. Asari, and Theus H. Aspiras. "Human detection in infrared imagery using intensity distribution, gradient and texture features." Mobile Multimedia/Image Processing, Security, and Applications 2018. Vol. 10668. International Society for Optics and Photonics, 2018.

[24] Khaldi, Belal, Oussama Aiadi, and Mohammed Lamine Kherfi. "Combining Color and Gray-Level Co-occurrence Matrix Features: A Comparative Study." IET Image Processing (2019).

[25] Korichi, Aicha, et al. "Off-line Arabic handwriting recognition system based on ML-LPQ and classifiers combination." 2018

International Conference on Signal, Image, Vision and their Applications (SIVA). IEEE, 2018.

[26] Wang, Yi, et al. "Polyp classification by Weber's Law as texture descriptor for clinical colonoscopy." Medical Imaging 2019: Computer-Aided Diagnosis. Vol. 10950. International Society for Optics and Photonics, 2019.

[27] Srivastava, Divya, et al. "Pattern-based image retrieval using GLCM." Neural Computing and Applications (2018): 1-14.