Semantic Based Image Retrieval through Combined Classifiers of Deep Neural Network and Wavelet Decomposition of Image Signal

Nadeem Qazi   B.L.Wlliam Wong

Department of Computer Science, Middlesex University, London, {n.qazi,w.wong} @mdx.ac.uk

Abstract

Semantic gap, high retrieval efficiency, and speed are important factors for content-based image retrieval system (CBIR). Recent research towards semantic gap reduction to improve the retrieval accuracy of CBIR is shifting towards machine learning methods, relevance feedback, object ontology etc. In this research study, we have put forward the idea that semantic gap can be reduced to improve the performance accuracy of image retrieval through a two-step process. It should be initiated with the identification of the semantic category of the query image in the first step, followed by retrieving of similar images from the identified semantic category in the second step. We have later demonstrated this idea through constructing a global feature vector using wavelet decomposition of color and texture information of the query image and then used feature vector to identify its semantic category. We have trained a stacked classifier consisting of deep neural network and logistic regression as base classifiers for identifying the semantic category of input image. The image retrieval process in the identified semantic category was achieved through gabor filter of the texture information of query image. This proposed algorithm has shown better precision rate of image retrieval than that of other researchers work

Keywords: image retrieval, wavelet decomposition, Gabor filter, semantic gap, stacked neural network

1 Introduction

Content-based image retrieval (CBIR) systems present a growing trend in all kind of applications including medicine, health care, internet, advertising, entertainment, remote sensing, digital libraries and crime detection. For example in medical domain, it is important to find similar images in various modalities acquired in various stages of disease progression to assist clinical decision-making process. Likewise, often during a criminal investigation, an analyst wishes to identify a digital piece of information such as unsubstantial images, tattoos, a criminal sketch generated from the details given by eyewitness or a crime scene from the huge database including both static and video images. Finding images that are perceptually similar to a query image is a challenging task in the dense database environments. There is a need for a better methodology for identifying the culprits from those images. Content-based image retrieval systems facilitate this process of image searching from large databases.

CBIR use the visual content of the query image to retrieve the best-matched image from the huge collection in the image database. The search is based on image features such as texture, shape, and color. A typical CBIR solution requires the construction of an image descriptor, which is characterized by (i) an extraction algorithm to encode image features into feature vectors; and (ii) a similarity measure to compare two images. The image retrieval system process is started by a user provided query image, followed by feature extraction of the image based on any appropriate feature selection method. A comparison of these features is then made with the features of the images in the database using some similarity measure. The matched images from the database are marked based on the value of the similarity measure, and the one having the highest value of the similarity measure is given to the user. Commonly used similarity distance measurements are Euclidean distance, Manhattan distance, Canberra distance matrix and histogram intersection distance. Researchers (Arevalillo-Herráez et al., 2008) have also suggested an algorithm to combine the similarity measurement distance based on the Bayes rule.

Previous studies on CBIR systems have focused on the global and local feature descriptor of the image. The commonly used visual descriptors are color, texture, shape and spatial relationship of the neighboring pixels in the picture. The feature extraction through a color descriptor depends on the selection of the appropriate color space. The commonly used color spaces are RGB, CIE CIE and HSV (or HSL, HSB).Color moment (Huang et al., 2010), color histogram (Sergyan, 2008) has also been used as feature descriptors in CBIR. However, a color descriptor for an image is not effective when there is a high spatial color variation. Thus, the researchers have investigated other low-level descriptors such as texture, which characterize the spatial distribution of gray levels in the pixel neighborhood.

The texture features of an image are identified by statistical, structural and spectral methods. The statistical methods for texture determination include power spectra, co-occurrence matrices, shift-invariant principal component analysis (SPCA), fractal model, and multi-resolution filtering techniques such as wavelet decomposition. However, (Selvarajah and Kodituwakku, 2011) have reported

DOI: 10.3384/ecp17142473

that the first order statistical method of determining image textures are less effective in image retrieval, with an average precision rate of 0.34. Their findings showed that the second order gray level co-occurrence matrix method performed better with an average precision rate of 0.44 . The same authors have also used the coefficients of Gabor, filtered as feature vectors to retrieve the images, and reported a precision rate of 0.76. (Zheng, 2015) proposed image retrieval system named as SIMPLIcity:(Semantics-sensitive Integrate Matching for Picture Libraries). They used histogram, color layout and coefficients of wavelet transform as feature vector over 600 medical images from six categories.

The feature descriptors, whether color, texture or shape, are all low-level features and are not able to truly exemplify the high-level concept in the user's mind. This problem is known as a semantic gap in the CBIR domain and is the main hindrance in the performance of the CBIR. Image annotation, Region- Based Image Retrieval (RBIR) approaches and relevance feedback have received more attention in recent years to overcome this gap. One of the approaches used to reduce the semantic gap is image annotation. The work presented in this paper, however, is based on class based annotation, representing the image retrieval as a multiple label classification problems where each class is defined as the group of database images labeled with a single semantic label.

We have proposed in this paper, that a categorical identification based on the semantic of the query image should first be established, before the retrieval and ranking process of the similar images from the identified semantic category of the given query image. Additionally based on the fact as reported by (Cleanu et al., 2007) that human eyes use of multi-scale linear decomposition for image texture, we used multi-resolution analysis techniques to extract the feature vectors of the query image. We constructed a global feature vector using both the color and texture information of the query image through its wavelet decomposition and used this feature vector to identify the semantic category of the query image. The local texture feature of the query image was then employed through gabor filtering to create a feature vector for retrieving and ranking the similar images from the identified class. Also in order to improve the accuracy of the identification of the semantic category of the image, we have utilized the combined classifier technique consisting of deep neural network and logistic regression. This approach was later compared with the previous research studies and was found to improve the precision rate of retrieved images. This approach thus may also be helpful in the research of reducing the semantic gap, assuming that the images are labeled into classes according to the semantics of the images.

The rest of the paper is structured as follows. We present the summary of related work in section 2 followed by proposed algorithm for image retrieval in section 3. Section 4 presents feature vector extraction based on daubechies wavelet decomposition and gabor filter followed by the algorithm for combining the classifiers for image semantic identification in section 5. We compare the performance result of our proposed algorithm in section 6 and paper is concluded in the concluding segment.

2 Related Work

(Hiremath and Pujari, 2007) have combined color and texture features using wavelet-based color histograms for image retrieval from the image databases of WANG. They have used the histogram intersection distance for determining the similarity between the query image and the database image. However, an image retrieval process defined in their work uses the algorithm to retrieve the images from the whole database without any identification of image category. The performance evaluation measurement precision for image retrieval for all the categories as reported in their paper falls between 7.2 and 7.5. (Wong et al., 2007) have used support vector machines and shape-based feature extraction for image classification.

Another emerging technique in CBIR is the use of deep learning neural network. (Krizhevsky et al., 2017) has used convolution neural network consisting of five convolution layers and pooling layers having 60 million parameters and 650,000 neurons to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. (Karande and Maral, 2013) have shown relevance feedback technique using artificial neural network trained feature vectors obtained from HSV model and texture to reduce the semantic though used the cloud computing to meet the challenge of computing power. (Wan et al., 2014) have investigated towards the effective role of deep learning in reducing the semantic gap their empirical study on Caltech256 dataset has revealed that pre-trained (convolutional neural networks) CNN model on large scale dataset are able to capture high semantic information in the raw pixels and can be directly used for features extraction in CBIR tasks. They, however, concluded that features extracted by pre-trained CNN model may or may not be better than the traditional hand-crafted features.

The research work presented in this paper is however, based on prior identification of the semantic category of input query image through a deep neural network, followed by retrieving closest similar image from the identified semantic category of the image. The detailed of the proposed algorithm is presented in the next section.

3 Methodology

The proposed algorithm of the CBIR as shown in Figure 1 uses the global descriptors to extract the characteristic of the image. The algorithm works in following steps:

- Create a global features space using wavelet decomposition of HSV color space of all the images in the database.

- Identify the semantic category of the query image. This is performed through training a combined classifier consisting of a Deep Neural Network and logistic regression under supervised, using the feature vector obtained in the first step.
- In the retrieving phase of image retrieval, extract global feature vector of the query image through gabor filter.
- Retrieve and rank similar images from the identified semantic category based on the euclidean distance of gabor feature vector of the query and similar candidate images.

3.1 DataSet

For testing our proposed algorithm we have used the image data set from the Wang image database and semantically divided into 10 categories such as horse, elephants, and beaches, dinosaurs, building, food, flowers, Africa, buses, and mountains. Each division was made of approximately 100 images of the same class. We divided these image data sets into two equal training and testing data set each consisted of 500 images and used them in training and testing the classifier.

4 Semantic Identification Through Multiple Classifiers

We then created a global feature vector space of all the images in the dataset. The feature vectors were extracted using the wavelet decomposition of 2D image signal discussed below:

4.1 Global Feature Extraction through Wavelet Decomposition

Wavelet transform plays a wide role in image processing and computer graphics due to its sub-band and multiresolution decomposition ability for describing the image features and characteristics and thus one of our reasons to use it for image decomposition and feature extraction. Discrete Wavelet Transformation (DWT) uses a short life mathematical wave function (t) as its base function known as wavelets to represent a continuous time signal into different scale components. The (Daubechies, 1988) has given the following mathematical equation of the Daubechies wavelet function (t) :

$$\Psi_{rjk}(x) = 2^{0.5j}\Psi_r(2^j x - k), j,k.r \in Z$$

Where j is scale,k is a translation and r is filter.

Due to Daubechies wavelet, efficiency in separating different frequency bands and reflecting all the changes in the neighboring pixels, we chose it to extract the feature vector image signal using (HSV) color space. The Daubechies wavelet filters were convoluted with each of the images in the database using two levels of resolution, generating high- and low-frequency bands of input images. We calculated this two-dimensional wavelet image transformation by computing row by row one-dimensional wavelet transformation in a horizontal direction, and then a column by column one-dimensional wavelet transformation in a vertical direction as shown in Figure 2. This produced the first level of decomposition. For the second level decomposition, we used this same process,however, using the low-level frequency component obtained in the first level decomposition. This finally yielded two levels of high and low-level frequency components generating four sub-images, which are labeled as LL, LH, HL and HH in the Figure 2, where

- Sub-image LL1 and LL2 represent the horizontal and vertical low-frequency part of the image at level 1 and 2 respectively and are recognized as an approximation.
- Sub-image HH1and HH2 represent the horizontal and vertical high-frequency part of the image at level 1 and 2 respectively and are called diagonal.
- Sub-image LH1 and LH2 represent the horizontal low and vertical high-frequency components at level 1 and 2 respectively and are known as horizontal.
- Sub-image HL1 and HL2 represent the horizontal high and vertical low-frequency components at level 1 and 2 respectively and are called vertical.

This two-level wavelet decomposition process generated approximation coefficients along with the detail coefficients in horizontal, vertical and diagonal direction at

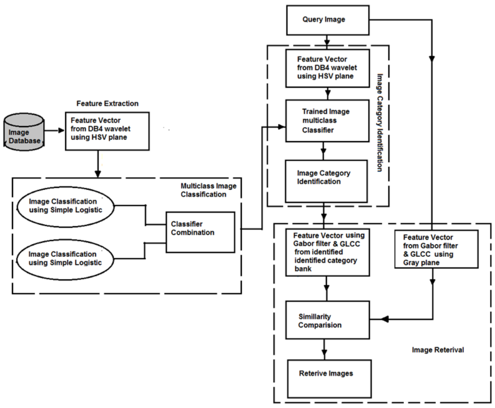
Query Image
Feature Vector from DB4 wavelet using HSV plane
Image Category Identification
Trained Image multiclass Classifier
Image Category Identification
Feature Extraction
Image Database
Feature Vector from DB4 wavelet using HSV plane
Multiclass Image Classification
Image Classification using Simple Logistic
Classifier Combination
Image Classification using Simple Logistic
Feature Vector using Gabor filter & GLCC from identified identified category bank
Feature Vector from Gabor filter & GLCC using Gray plane
Simillarity Comparision
Image Reterival
Reterive Images

Figure 1. Proposed algorithm of the CBIR.

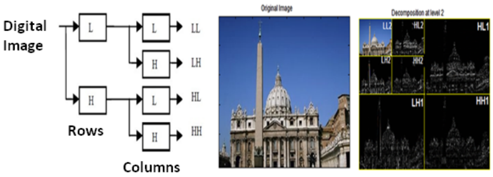

Figure 2. Wavelet Decomposition of 2D-image signal.

DOI: 10.3384/ecp17142473

each level. We then constructed an image feature vector consisting of 33 elements for each of the images in the database. Each image feature vector consisted of standard deviation, mean, skewness and kurtosis of the histogram of the detail coefficients and energy vector of both the approximation and detail coefficients obtained in the two-level decomposition of the input signal. This image feature vector was later used to identify the semantic class of the query image.

4.2 Deep Learning of Neural Network

We then used the extracted global feature vector to train classifiers for identification of the semantic category of the query image. To achieve this we defined an image domain space consisting of N samples of images, each represented by a global feature vector q (f1, f2, f3....f33) to be used as input feature vector to the classifier. The goal was set to assign every input query image a semantic class label Ci from the class labels space C (C1,..C10) using these trained classifiers. We employed the deep learning approach in training the classifier to identify the semantic category of the query image. Two classifiers were explored i.e. deep neural network and logistic regression. Deep learning refers to a class of machine learning techniques that employ deep architecture, unlike their shallow counterpart processing information through multiple stages of transformation and representation. Deep learning neural network architectures are different from "normal" neural networks because they have more hidden layers.

A Deep Neural Network consisting of one input layer, three hidden layers, and one output layer each having sigmoid activation function was constructed. The choice of the activation function was made following the research work of (Shenouda, 2006). They have performed a quantitative comparison of the four most commonly used activation functions, including the Gaussian RBF network, over ten real different datasets to show that the sigmoid activation function is a substantially better activation than others. Back propagation training algorithm was employed, to 10 fold classification. Feature vectors generated through the process of Daubechies wavelet decomposition were used as external inputs to five layers deep learning neural networks during the training phase. Following equation was used to calculate network output after each layer of the selected neural architecture.

$$a^{(i+1)} = S^{(i+1)}(W^{(i+1)}a^i + b^{(i+1)})$$

Where i=1,2,3,4
a,, is the output from i^th neural network layer
$S^{(i+1)}$, is the sigmoid function
$W^{(i+1)}$ and $b^{(i+1)}$ are neuron weights The input a^0, consisting of feature vector $f_{1,1}$, to $f_{n,33}$, to the input layer is then given as:

$$a^0 = q = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ x_{d1} & x_{d2} & x_{d3} & \dots & x_{dn} \end{bmatrix}$$

The network outputs generated by the output layer of the network are given by

$$a^4 = C = \begin{bmatrix} C_1 & C_2 & C_3 & \dots & C_k \end{bmatrix}$$

Where n is the number of observations and the training set in the form $(q_1\ C_1), (q_1\ C_1), (q_n\ C_k)$ were presented to the feed forward neural network during training.

In addition to these, we also chose logistic regression as of our second classifier. However, in this case, we estimated a probability of a given instance Cj belonging to semantic class space C=(C1,..C10). The probability for class j excluding the last class for multi-class problems was determined by

$$p_j(f_j) = \frac{e^{f_i\theta_j}}{\sum_{j=1}^{k-1}(1+e^{f_i\theta_j})}$$

The probability of the last class was calculated using the following equation

$$1 - \sum_{j=1}^{k-1} p_j(f_j) = \frac{e^{f_i\theta_j}}{\sum_{j=1}^{k-1}(1+e^{f_i\theta_j})}$$

Where k is the number of classes, n is the total number of observations, f_i is the input feature vector and $theta$is the parameter matrix, which is calculated using the Quasi-Newton Method.

The two classifiers were then fused to improve the classification accuracy. This approach of classification through fusion is increasingly embraced by researchers in recent years. (Baskaran et al., 2004) have combined weighted multiple classifiers consisting of naive Bayes, artificial neural networks, fuzzy C-mean classifier and variants of distance classifiers for the remote sensing image classification. (Qazi and Raza, 2012) has suggested using combined classifiers for the better classification of network intrusion minor classes.

We used the stacking method to combine the trained classifiers. The stacking of classifiers algorithm is relatively a new approach in classifier combination and consists of classifiers at two levels i.e. base classifiers and Meta classifier or arbiter. The Meta classifier selects the best classifier among several base classifiers. We used linear regression as the training algorithm for the Meta learner to stack the two base classifiers i.e. deep Learning Neural Network and logistic regression. The training dataset was then used to train this fused classifier for identification of semantic image category.

5 Image Retrieval

The retrieval phase of the similar images from the matched semantic category uses textural features of the input query image. We constructed texture feature vector of the images using the localized feature of the image through Gabor filter due to its wavelet nature capturing energy at a specific frequency and a specific direction. The mathematical representation of a 2-D Gabor filter (Gb), having wavelength represented by λ, orientation angle by $theta$, and standard deviation along x and y by $\sigma_x$and $\sigma_y$respectively, may be given by the following relation:

$$\text{Gb(x,y)=}\ Gs(x,y)e^{j\lambda(xcos\theta+ysin\theta)}$$

Where Gs(x,y) is the Gaussian function given by

$$Gs(x,y)=\frac{e^{-0.5((x/\sigma_x)^2+(y/\sigma_y)^2)}}{\sqrt{2\pi\sigma_x\sigma_y}}$$

We generated a filter bank consisting of 36 filters by varying the wavelength λ from 2.5 to 3.0 with an increment of 0.1, and the orientation angle $\theta from \pi, 3/2\pi, 11/6\pi, 25/12\pi, 187/60\pi, and\ 441/180\pi$. Each of the filters in the filter bank was then convoluted with the input image finally the feature vector consisting of 144 elements was constructed by calculating the contrast, homogeneity, correlation and energy of the GLCM (Gray Level Co-occurrence Matrix) for each filtered image in the filter bank.

6 Performance Analysis

In order to assess the performance of the proposed algorithm, we used the test data set consisting of 250 images approximately 25 from each category. For each tested query image two feature vector was extracted , wavelet based feature vector was used to identify the semantic category and then another Gabor filter-based feature vector of the query image was used to retrieve the smaller images from the identified category using euclidean distance.

The precision and recall rate of the classification obtained by the combined classifier of deep neural network and logistic regression for the testing dataset is shown in Table 1. However, Figure 3 shows precision and recall rate for all the semantic categories used in the test data. Some of the visual queries and retrieved similar images are presented in Figure 4. The average retrieval rate was found to be less than 1 minute approx 40 to 50 seconds for retrieving four similar images of the query image. A precision rate of all the retrieved images using a test data set has been calculated and shown in Table 2 in comparison with the algorithms of other researchers. It can be seen that our algorithm has performed slightly better it is because that the identifying the right category of query image increases the probability of the retrieving similar image.

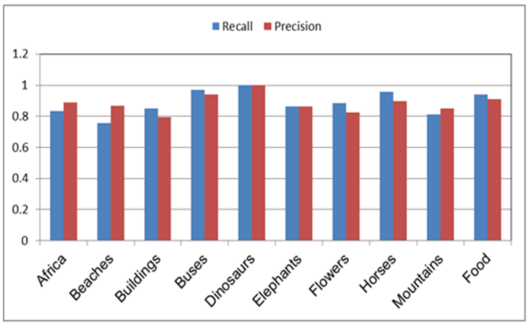

Figure 3. Classification accuracy of the combined classifier.

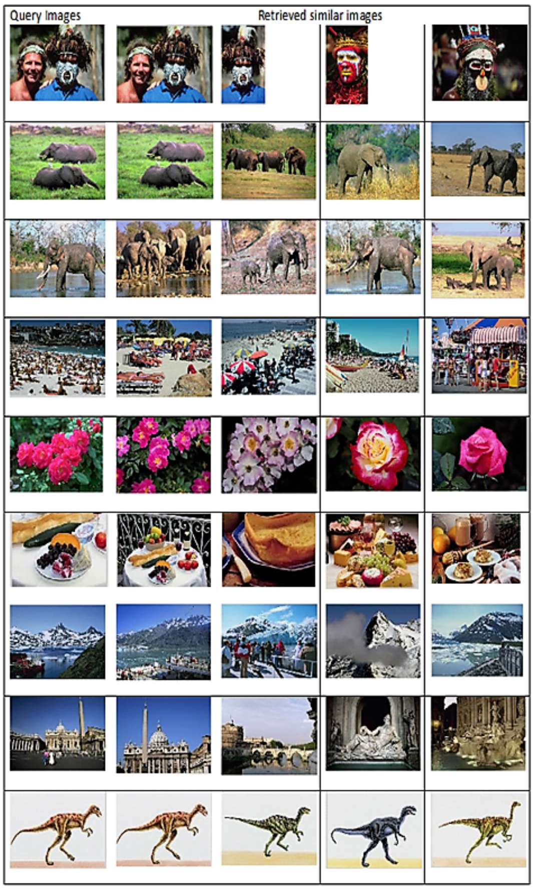

Figure 4. Result of visual queries through the proposed algorithm.

Table 1. Trained classifiers accuracy.

Instances	ANN	Logistics	Stacked Classifiers
Correctly Classified	83.5%	85.9%	87.0%
Incorrectly Classified	16.5%	14.1%	13.0%

7 Conclusions

It was noted in this research that classification of the images in the semantic based categories classes may be help-

Table 2. Performance Comparison With Other Algorithms.

Category	Proposed algorithm %	Ahmed J.Afifi Wasam Ashour %	Manimala Singha et al %	Chen-Horng Lin et al %	M Babu et al %
Africa	0.74	0.71	0.65	0.68	0.56
Beaches	0.90	0.85	0.62	0.54	0.53
Buildings	0.86	0.83	0.71	0.56	0.6
Buses	0.82	0.85	0.92	0.89	0.89
Dinosaurs	0.99	0.99	0.97	0.99	0.98
Elephants	0.76	0.71	0.86	0.66	0.57
Flowers	0.94	0.93	0.76	0.89	0.89
Horses	0.89	0.57	0.87	0.80	0.78
Mountains	0.86	0.42	0.49	0.52	0.51
Food	0.85	0.97	0.77	0.73	0.69
Average	0.86	0.78	0.76	0.72	0.70

ful in reducing the semantic gap. It also improves the retrieval efficiency because after the related semantic class of input query image is identified then retrieval of similar images is performed within the group of more related images in the same class. We aim to apply the proposed technique in developing a robust image and video search engine that could assist the analyst in retrieving photographs, images of the criminals or crime scenes from huge criminal database such as VALCRI database.

Acknowledgment

The research leading to the results reported here has received funding from the European Union Seventh Framework Program through Project VALCRI, European Commission Grant Agreement NÂř FP7-IP-608142, awarded to Middlesex University and partners

References

Miguel Arevalillo-Herráez, Juan Domingo, and Francesc J Ferri. Combining similarity measures in content-based image retrieval. *Pattern Recognition Letters*, 29(16):2174–2181, 2008.

R Baskaran, M Deivamani, and A Kannan. A multi agent approach for texture based classification and retrieval (matbcr) using binary decision tree. *International Journal of Computing & Information Sciences*, 2(1):13, 2004.

Ctlin Cleanu, De-Shuang Huang, Vasile Gui, Virgil Tiponu, and Valentin Maranescu. Interest operator versus gabor filtering for facial imagery classification. *Pattern Recogn. Lett.*, 28(8):950–956, June 2007. ISSN 0167-8655. doi:10.1016/j.patrec.2006.12.013. URL `http://dx.doi.org/10.1016/j.patrec.2006.12.013`.

P. S. Hiremath and J. Pujari. Content based image retrieval using color, texture and shape features. In *15th International Conference on Advanced Computing and Communications (ADCOM 2007)*, pages 780–784, Dec 2007. doi:10.1109/ADCOM.2007.21.

Z. C. Huang, P. P. K. Chan, W. W. Y. Ng, and D. S. Yeung. Content-based image retrieval using color moment and gabor texture feature. In *2010 International Conference on Machine Learning and Cybernetics*, volume 2, pages 719–724, July 2010. doi:10.1109/ICMLC.2010.5580566.

S. J. Karande and V. Maral. Semantic content based image retrieval technique using cloud computing. In *2013 IEEE International Conference on Computational Intelligence and Computing Research*, pages 1–4, Dec 2013. doi:10.1109/ICCIC.2013.6724277.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, May 2017. ISSN 0001-0782. doi:10.1145/3065386. URL `http://doi.acm.org/10.1145/3065386`.

N. Qazi and K. Raza. Effect of feature selection, smote and under sampling on class imbalance classification. In *2012 UKSim 14th International Conference on Computer Modelling and Simulation*, pages 145–150, March 2012. doi:10.1109/UKSim.2012.116.

S Selvarajah and SR Kodituwakku. Analysis and comparison of texture features for content based image retrieval. 2011.

S. Sergyan. Color histogram features based image classification in content-based image retrieval systems. In *2008 6th International Symposium on Applied Machine Intelligence and Informatics*, pages 221–224, Jan 2008.

Emad A. M. Andrews Shenouda. A quantitative comparison of different MLP activation functions in classification. In *Proceedings of the Third International Conference on Advances in Neural Networks - Volume Part I*, ISNN'06, pages 849–857, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 3-540-34439-X, 978-3-540-34439-1. doi:10.1007/11759966_125. URL `http://dx.doi.org/10.1007/11759966_125`.

Ji Wan, Dayong Wang, Steven Chu Hong Hoi, Pengcheng Wu, Jianke Zhu, Yongdong Zhang, and Jintao Li. Deep learning for content-based image retreval: A comprehensive study. In *Proceedings of the 22nd ACM International Conference on Multimedia*, MM '14, pages 157–166, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3063-3. doi:10.1145/2647868.2654948. URL `http://doi.acm.org/10.1145/2647868.2654948`.

Wai-Tak Wong, Frank Y. Shih, and Jung Liu. Shape-based image retrieval using support vector machines, Fourier descriptors and self-organizing maps. *Information Sciences*, 177(8):1878 – 1891, 2007. ISSN 0020-0255. doi:https://doi.org/10.1016/j.ins.2006.10.008. `http://www.sciencedirect.com/science/article/pii/S0020025506003227`.

K. Zheng. Content-based image retrieval for medical image. In *2015 11th International Conference on Computational Intelligence and Security (CIS)*, pages 219–222, Dec 2015. doi:10.1109/CIS.2015.61.