

Experimental Analysis of Medical Image Classification and Retrieval Techniques

P. Nalini^{1*}, B. L. Malleswari²

¹Assistant Professor, ECE, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, India.

Email: pasumarthinalini@gmail.com

²Principal, Sri Devi Women's Engineering College, Hyderabad, Telangana, India.

Email: blmalleswari@gmail.com

*Corresponding Author

Abstract: Medical Image classification and similar image retrieval are the two important processes in diagnosis and automatic annotation. These help the doctors and radiologists in their decision making during disease identification and decision making. Image classification is usually done by checking its content similarity. Image content is its visual features referring to mathematical attributes. Similarity checking is done by using similarity or dissimilarity measures which are also known as distance metrics. As image attributes are wide in range, the similarity measure worked well for one feature set may not show the similar performance for other. For this reason in this paper we explored various existing similarity measures viz. Manhattan, Cosine, Chi-square and Cramer distances and their effect with respect to image intensity features and wavelet based texture features. We drawn certain conclusions on the performance of these distance metrics in classification and retrieval of IRMA data sets. Mean Average Precision and Average Recall Rates are used in analyzing retrieval performance for analyzing the medical image retrieval and classification task.

Keywords: Classifiers, Distance metrics, Medical images, Retrieval.

I. INTRODUCTION

Image attributes mostly classified as Color, texture and shape. Most of the medical imaging modalities such as XRAY, CT and MRI produced medical images are usually in gray tone. Therefore color features like color histograms and color correlograms would not work well in analyzing visual content in medical images. Texture features are more powerful features in analyzing the medical images. Knowledge extractable from medical images is not precise. Spatial data in the image is not expressed in conventional languages. Most of the image has geometrical information and Medical images arising

from photography (e.g., endoscopy, histology, dermatology), radiographic projection (e.g., x-rays, some nuclear medicine), and tomography (e.g., CT, MRI, ultrasound) impose unique, image-dependent restrictions on the nature of features available for abstraction. Similarity from a medical perspective is predominantly context dependent [6].

Different modality based medical images have different characteristics such as Microscopic histology images possess unique color signatures and cell textures. Ultrasound images of large organs appear to be dominated by textures, hence emphasize extracting a global property rather than local features, Image arising from a projection technique, from a tomographic technique such as MRI, Tomographic images grouped by acquisition from individual subjects also have the unique virtue of retaining the data required for an ambiguous 3-D reconstruction of tissue structures Shape matching similarity operations are used here.

Chest X ray images are projections of many overlapping structures. Indexing procedures address textures based features rather shape. Tomographic images readily permit non overlapping geometrically bounded organs and tissues to be identified as a collection of individual features [6]. Texture contains important information about the structural arrangements of their surfaces and their relationships to the surrounding environment. Small area of patches with little gray variation is called TONE. Small area of patches with large gray variation is called TEXTURE. GLCM Features computed for various angular relationships with the distance between pixels. Four 4 x 4 matrices are generated as $f(i,j,d,0^0)$, $f(i,j,d,45^0)$, $f(i,j,d,90^0)$, $f(i,j,d,135^0)$. 'd' varies from 1:n. [1]. Later Tamura *et al.* approximated six textural features contrast, coarseness, directionality, line likeliness, regularity and roughness, which works similar to psycho visual (human visual) perception. Contrast, Coarseness and directionality considered as most powerful features among all texture attributes of an image discussed in [2]. Texture classification and discrimination based on the energies of image subbands using DCT, wavelet and spatial partitioning detailed in [3].

The wavelet transform involves filtering and sub sampling. Compared non orthogonal (Gabor), orthogonal and biorthogonal (tree structured) decompositions wavelet transform decompositions to analyze texture classification [4]. Mean and standard deviation of Daubechies based wavelet coefficients are used as features. Quad tree indexing, Energy estimation (mean and variance) in subbands are used for feature extraction from subbands. Images decomposed into blocks and wavelet coefficients are computed for each block and the query features compared with these features for fast retrieval of similar images from large databases discussed in [5].

Transformation based texture calculation based on energy of subbands using DCT presented in [15]. Statistical textural features obtained with the help of discrete wavelet transforms presented in [16]. Gabor wavelet texture features compared with orthogonal wavelet texture features in [17]. In Operator or pattern based texture calculations, Local Binary Patterns (LBP), Local Ternary Patterns (LTP) and Local Derivative Patterns (LDP) are used. These are more powerful texture features as they are invariant to rotation and scaling proposed in [18].

Color and texture features were compared with nine different similarity measures including Heuristic histogram distances (Minkowski form, Histogram intersection, Weighted mean variance), Non parametric statistical form (Kolmogorov Smirnov, Cramer von Mises, Chi-square), Information theoretic divergences (Kullback - Leibler divergence, Jeffery divergence), ground distance measures (Earthmover's and quadratic distance) in [7], [10]. The selection of a similarity measure substantially improves the efficiency of classifier or retrieval. Histogram quadratic distance incorporates cross bin information via a similarity matrix A. Earth Movers Distance computations complexity is highest. It is minimum cost of transferring one distribution with the other. EMD can be defined as a solution of the transportation problem. Jeffrey divergence is the symmetric version of KL divergence. Chi square measures line likeliness from one distribution being drawn from the other. Perform better with large number of observations. Cramer von Mises used for judging the goodness of a fit of a cumulative distribution functions when compared with empirical distribution function. Information Theoretic (How compact one distribution can be coded using the other one as codebook) Kullback Leibler Divergence (KLD) and Jeffrey Divergence (JD) are the examples of information theoretic. Statistical distances used to test the hypothesis that two empirical distributions that have been generated from the same underlying true distribution. These include Kolmogorov-Smirnov Divergence (KSD), Cramer-von Mises (CvM), Chi Square and Pearson Correlation Coefficient. Ground based distances perceptually meaningful distance measures between individual features. These include Quadratic distance and Earth Mover's Distance (EMD). Manhattan is best among Minkowski distance metrics. JD is stable over KLD. For large samples chi-square and statistical distances worked better [7], [10].

II. IRMA DATABASE

IRMA image data sets and retrieval of similar modality images based on prior learning of classifier presented by RWTH Aachen university developers, Mono hierarchical multi axial classification code is presented in [8]. 14 digit code is used in IRMA database with first five digits represent TECHNICAL CODE relating medical imaging modality with technical parameters in which digit 1 indicate imaging physical technique (X-ray, US, CT), digit 2 represent modality position (plain projection, fluoroscopy, angiography etc), digit 3 denote technique (digital, analog etc), digit 4 for assessing sub techniques (high energy, low energy, parallel beam, etc) and digit 5 for external ads, drugs and additional markers. Next three digits indicate directional code for orientation of the image with respect to body Digit1: common orientation (coronal, sagittal, transversal etc). Digit 2: specific orientation (posterior - anterior, etc) and Digit 3: functional orientation (standing, sitting, inclination) etc. Next 3 digits ANATOMICAL CODE: for body region examined Nine major anatomical regions are extracted. (1. Total body, head / skull, spine etc) and last 3 digits: BIOLOGICAL CODE for biological system under evaluation. (The top level code ten organ systems are specified like cerebral spinal, cardiovascular, respiratory, and gastrointestinal and so on. The code axis is orthogonal and each axis is built mono hierarchically [8].

Assert system analysis is CT images of the lung with respect to 8 certain diagnostic inquires KMeD and COBRA retrieve ventricular shapes extracted from MR images of the head.

I-Browse operate on histological slices Interpretation of medical images is dependent on both image and query context.

I-Browse is a CBIR system that integrates iconic and semantic features for histological image analysis and also to do textual annotations for unknown images. Coarse feature detectors use color and gray level histograms and semi-fine feature detection is done by Gabor filter [9].

IRMA concept is based on a conceptual and algorithmic separation of seven processing steps: 1. Categorization using global features 2. Determination of parameters for registration geometry and contrast for each likely category. 3. Feature extraction using local features, Feature selection and combination with respect to category and query content. 4. Indexing resulting in a hierarchical multi scale blob representation and registration. 5. Identification of blobs linking a priori knowledge to image content. 6. Identification of blobs by finding priori knowledge to image content. 7. Retrieval processed on the abstract blob level [8].

III. DISTANCE MEASURES

Minkowski distance metric at different levels presented in [14]. These metrics are preferred when each dimension holds equal

importance in retrieval process. Minkowski metric was used for feature vector comparison in [19]. Manhattan distance or City block distance or Minkowski L_1 depends on the rotation of the coordinate system. Cosine Angle Distance (CAD), Chi-square and Mahalanobis distance measures for shape databases evaluated in [20]. According to [21] and [22] the discrete, three dimensional analogous of the Kolmogorov-Smirnov test L_∞ , Cramer -von Mises test (L_2) and the Earth Mover Distance (L_1) are the statistical distance metrics.

A. Minkowski Distance

The Minkowski distance of order p between two points X and Y belongs to feature space R , $X = (x_1, x_2, x_3, \dots, x_n)$ and $Y = (y_1, y_2, y_3, \dots, y_n) \in R$ is given in equation (1).

$$d(x, y) = \sum_{i=1}^n (|x_i - y_i|^p)^{\frac{1}{p}} \quad (1)$$

In which p represent the order of the distance metric. When $p < 1$, it is not considered as a distance metric as it does not satisfy the triangular inequality. When $p = 1$, the distance turns into Manhattan distance and when $p = 2$ it become Euclidean distance. In the limiting case of p reaching infinity, the Chebyshev distance is obtained. Minkowski distance metric is generalized form of Euclidean distance and city block distances, preferred when each dimension holds equal importance in retrieval process.

B. Manhattan or City Block Distance

This metric0 measures the direct grid distance along the pixels and diagonal movements not allowed. Manhattan distance metric retrieve images at a faster rate when compared with Euclidean distance [10]. The metric shown good MAP in both feature similarity measures, but worked well for gray histogram comparison over texture feature similarity. This distance is shown in equation (2).

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

C. Euclidean Distance

This distance represent length of the line segment connecting two points in a feature set. In image processing and retrieval, the images are with n dimensional feature vectors and hence n - dimensional Euclidean distance is used. If x and y represent two images then the Euclidean distance between them obtained as shown in equation (3).

$$d(x, y) = \sum_{i=1}^n \sqrt{(x_i - y_i)^2} \quad (3)$$

In this paper, we used the Euclidean distance for gray intensity and texture feature comparisons.

D. Chebyshev / Chessboard or Infinity Distance

This distance also known as chessboard obtained when limiting value reaches to infinity. This distance between two points (x, y) is expressed as shown in equation (4).

$$d(x, y) = \max_{i=1}^n |x_i - y_i| \quad (4)$$

E. Cosine Angle Distance

Cosine Angle Distance (CAD) does not follow the triangular similarity. The cosine distance metric normalizes all feature vectors to unit length and makes it invariant against relative in-plane scaling transformation of the image content. This measure is best suited to find the orthogonality between two vectors. If the cosine angle computed between the Eigen values of vectors, it works much better. CAD is shown in equation (5).

$$d(x, y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}} = \frac{x_i \cdot y_i}{\|x_i\| \|y_i\|} \quad (5)$$

F. Chi-Square Distance

The chi-square distance measure is used in correspondence analysis and related ordination techniques. Chi-squared distance does not reach a constant, maximal value for sample pairs with no species in common, but fluctuates according to variations in the representation of species with high or low total abundances. Chi-squared tests are often constructed from a sum of squared errors, or through the sample variance as shown in equation (6).

$$d_{\text{chisquare}}(x, y) = \sum_{i=1}^N \frac{(x_i - \mu_i)^2}{\mu_i} \quad (6)$$

G. Kullback-Leibler Distance

Solomon Kullback and Richard Leibler introduced Kullback Leibler divergence in 1951 [38]. It does not obey triangular inequality hence it is not a valid distance metric. This also considered as relative entropy of two distributions and is shown in equation (7).

$$d_{\text{kullback}}(x, y) = \sum_{i=1}^N x_i \log \frac{y_i}{x_i} \quad (7)$$

H. Jeffrey Distance

Jeffrey Divergence is the symmetric version of Kullback-Leibler distance with respect to samples x and y . The divergence equation is shown in equation (8).

$$d_{\text{jeffray}}(x, y) = \sum_{i=1}^N x_i \log \frac{y_i}{\mu_i} + y_i \log \frac{x_i}{\mu_i} \quad (8)$$

I. Kolmogorov Smirnov Divergence

This is a non-parametric test and calculates distance between empirical distribution function of the sample with cumulative distribution function of the reference. KS distance shown in equation (9).

$$D_{L_\infty}(x, y) = \max_{i=1}^N |(X_n - Y_n)| \quad (9)$$

J. Cramer Von Mises Divergence

This metric tests the goodness of fit for cumulative distribution function and to compare two empirical distributions. Harald Cramer and Richard Edler von Mises propose this metric [12] [13]. It is an alternative for Kolmogorov Smirnov test.

$$D_{L_1}(x, y) = \sum_{i=1}^N |(X_i - Y_i)| \quad (10)$$

K. Earth Mover's Distance

It is a measure of distance between two probability distributions. It works to find the minimum distance function. EMD is shown in equation (11).

$$D_{L_2}(x, y) = \sum_{i=1}^N \sqrt{(X_i - Y_i)^2} \quad (11)$$

IV. METHODOLOGY

In this paper we presented image retrieval task based on gray histogram features and Haar wavelet based texture features for analyzing the performance of standard geometrical, statistical and cumulative distance measures consisting Euclidean, Manhattan, Chebyshev, Cosine angle, Chi-square, Kullback-Leibler, Jeffrey, Kolmogorov-Smirnov, Cramer von Mises and Earthmover's distances. Experimentation is done in Matlab to analyze the effect of distance metrics on different types of images.

1. Image database is loaded into Matlab workspace.
2. Query Image is selected from the database.
3. Intensity based feature extraction is done by computing gray histograms for query and database images.
4. Using Discrete Haar Wavelet transform (DWT), texture features extracted.
5. Geometrical distance measure Manhattan, statistical distance metrics Chi-square, cumulative statistical measure Cramer and Cosine angle distance were applied for feature similarity measurement.
6. Performance measures, precision (P) as given in eq.13, Mean average precision MAP and recall (R) as given in eq.12. are evaluated for the retrieved images.
7. Comparative analysis is done for image retrieval based on effect of similarity measures for gray intensity and texture features.

Precision and recall are used as measurements of classification in CBIR of medical images which are defined as:

$$\text{Recall} = \frac{\text{Total number of relevant images retrieved}}{\text{Total relevant images in the data base}} \quad (12)$$

$$\text{Precision} = \frac{\text{Total number of relevant images retrieved}}{\text{Total number of retrieved images}} \quad (13)$$

V. EXPERIMENTS AND RESULTS

We did experimentation of our method on IRMA 2007 and 2008 datasets for classification and retrieval of twenty different groups of anatomical images including lungs, spine, hands, legs ankle, arms and skull images. Sample query images are shown in Fig. 1. We presented the comparative analysis of which distance metric performed well for a particular type of query for its gray and texture features comparison. Gray features computed from 64-bin histogram and Discrete Wavelet coefficients computed to represent texture features.

In this paper we experimented on four distance measures Manhattan, Cosine Similarity, Chi-square and Cramer chosen each from the category of geometrical distances, statistical distances and cumulative distances. We compared these distances in terms of Mean Average Precision (MAP) and Average Retrieval Rate (ARR). Among these distances Manhattan distance and Cosine angle distances were shown outstanding performance in terms of MAP and ARR. We also analyzed that wavelet coefficients representation of image features worked in much better way in image classification as well as retrieval when compared with computation of conventional gray scale histograms to represent images.



Fig. 1: Sample Query Images

We also understood from the analysis that medical anatomical images are rich in texture details and the feature set that worked well for one anatomical structure may not show the same effect on other. As an example from Table I it is clearly shown that

for knee and skull images cosine angle distance shown superior performance where as for lungs, spine, neck and hand images Manhattan shown better retrieval compared to others. In terms feature sets, for feet and knee images intensity features worked well over texture features in classification and retrieval. We selected 1000 images from IRMA dataset consisting 50 images of twenty different categories. Selected 5 images of each category as queries and tested the system. Average precision and Recall rate computed for retrieval up to 50 images in terms of each distance metric.

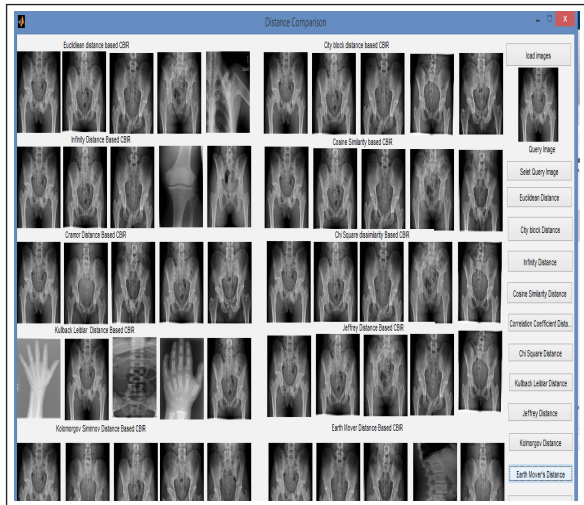


Fig. 2: Abdomen Images with Gray Intensity Features

Through this experimentation it is observed that Manhattan and Cosine angle distances performed well with DWT features with among all other distance metrics as shown in Table I. Image retrieval performed using gray histograms shown in Fig. 2. Retrieval of spine and lung images using wavelet texture features shown in Fig. 3 and Fig. 4.

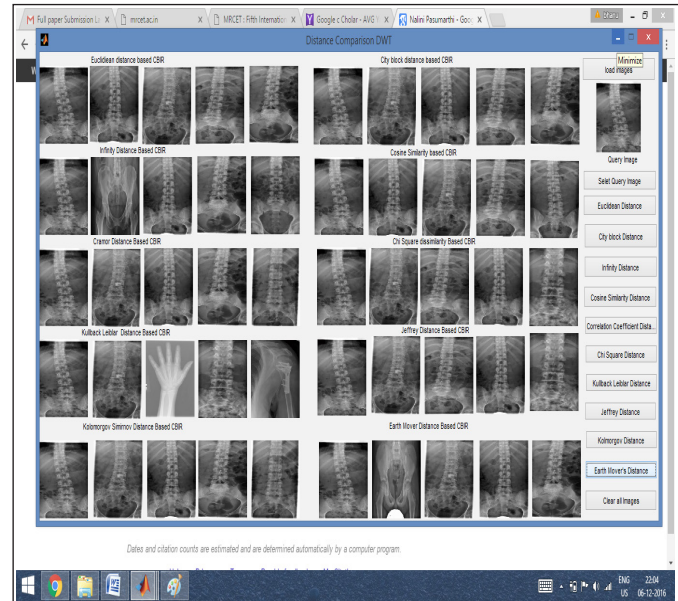


Fig. 3: Spine Images with DWT Texture Features

TABLE I: PRECISION AND MAP AND ARR FOR INTENSITY AND WAVELET FEATURES

Class of Images	Manhattan		Cosine Angle		Chi Square		Cramer von Mises	
	DWT	HG	DWT	HG	DWT	HG	DWT	HG
Abdomen	0.96	0.58	0.86	0.51	0.87	0.63	0.37	0.46
Spine	0.80	0.71	0.77	0.66	0.80	0.7	0.31	0.54
Lungs (front view)	0.93	0.49	0.95	0.48	0.83	0.48	0.45	0.45
Lungs (side View)	0.94	0.86	0.86	0.56	0.87	0.86	0.38	0.68
Mammogram	0.67	0.35	0.5	0.35	0.52	0.33	0.63	0.45
Neck	0.36	0.36	0.3	0.36	0.28	0.22	0.1	0.29
Hand Fingers	0.38	0.24	0.35	0.26	0.31	0.29	0.19	0.27
Wrist	0.50	0.49	0.42	0.28	0.42	0.45	0.2	0.34
Hands	0.45	0.12	0.46	0.11	0.29	0.12	0.12	0.11
Elbow	0.39	0.17	0.33	0.18	0.35	0.15	0.11	0.15
Feet	0.49	0.39	0.41	0.29	0.34	0.31	0.08	0.29
Ankle	0.38	0.37	0.28	0.36	0.35	0.4	0.15	0.34
Knee (side view)	0.31	0.44	0.33	0.32	0.71	0.43	0.18	0.33
Knee (front view)	0.73	0.34	0.71	0.29	0.59	0.34	0.16	0.29
Knee (top view)	0.57	0.62	0.76	0.63	0.39	0.64	0.1	0.37
Shoulder	0.87	0.29	0.5	0.18	0.88	0.33	0.19	0.22
Skull (front view)	0.48	0.69	0.89	0.55	0.38	0.73	0.32	0.46
Skull (Side view)	0.65	0.56	0.37	0.46	0.58	0.51	0.11	0.42

Skull (top view)	0.39	0.37	0.47	0.29	0.36	0.39	0.23	0.29
Teeth	0.45	0.35	0.46	0.28	0.23	0.35	0.14	0.27
MAP	0.59	0.44	0.55	0.37	0.52	0.43	0.23	0.35

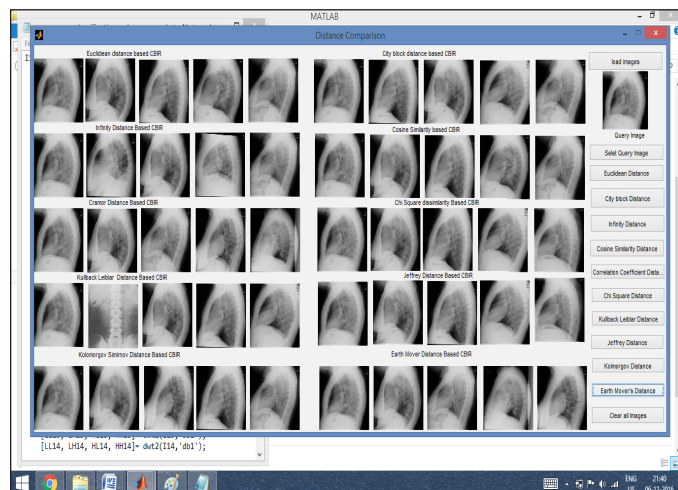


Fig. 4: Lung Images with DWT Texture Features

VI. CONCLUSIONS

In this paper, first we presented an overview of geometric and statistical distance metrics used in image retrieval applications and performed the comparative analysis of four diversified distance measures including Manhattan, Cosine Angle, Chi-square, Cramer von Mises and Earthmover's distance on intensity and texture features. Intensity features extracted by computing gray level 64-bin histograms and texture features through wavelet decompositions. Among geometrical distances Manhattan distance shown outstanding performance and in statistical distance metrics Cosine Similarity, shown good MAP score for all the types of queries. Finally it is observed that geometrical and statistical distance measures - Manhattan, Cosine Angle shown good MAP. Through this experimentation we draw a conclusion that medical image retrieval and classification not only depends on the image content representation but also on the similarity / distance measures used.

REFERENCES

- [1] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions*, pp. 610-621, 6th Nov. 1973.
- [2] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *Systems, Man and Cybernetics, IEEE Transactions*, pp. 460-473, 8th June, 1978.
- [3] J. R. Smith, and S.-F. Chang, "Transform features for texture classification and discrimination in large image databases," *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, vol. 3. IEEE, 1994.
- [4] W. Y. Ma, and B. S. Manjunath, "A comparison of wavelet transform features for texture image annotation," *Second International Conference on Image Processing (ICIP '95)*, Washington, D.C., vol. 2, pp. 256-259, Nov. 1995.
- [5] E. Remias, G. Sheikholeslami, A. Zhang, and T. S. Mahmood, "Supporting content-based retrieval in large image database systems," *Multimedia Database Management Systems*, Springer US, pp. 61-78, 1997.
- [6] H. D. Tagare, C. C. Jaffe, and J. Duncan, "Medical image databases," *Journal of the American Medical Informatics Association*, vol. 4, no. 3, pp. 184-198, 1997.
- [7] J. Puzicha, J. M. Buhmann, Y. Rubner, and C. Tomasi, "Empirical evaluation of dissimilarity measures for color and texture," In *Computer Vision, Proceedings of the Seventh IEEE International Conference*, vol. 2, pp. 1165-1172, 1999.
- [8] T. M. Lehmann, B. B. Wein, D. Keysers, M. Kohnen, and H. Schubert, "A monohierarchical multiaxial classification code for medical images in content-based retrieval," In *Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium on*, pp. 313-316. Jul. 2002.
- [9] L. H. Y. Tang, R. Hanka, and H. H. S. Ip. "Histological image retrieval based on semantic content analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 1, pp. 26-36, 2003.
- [10] M. P. Nalini, and B. L. Malleswari, "An empirical study and comparative analysis of Content Based Image Retrieval (CBIR) techniques with various similarity measures," *3rd International Conference on Electrical, Electronics, Engineering Trends, Communication, Optimization and Sciences (EEECOS)*, pp. 373-379, June 2016.
- [11] M. P. Nalini, and B. L. Malleswari, "Review on content based image retrieval: From its origin to the new age," *International Journal of Research Studies in Science, Engineering and Technology*, vol. 3, no. 2, pp. 18-41, Feb. 2016.
- [12] H. Cramér, "On the composition of elementary errors," *Scandinavian Actuarial Journal*, no. 1, pp. 13-74, 1928.
- [13] R. E. Von Mises, "Wahrscheinlichkeit, statistik und wahrheit," Julius Springer, 1928.
- [14] F. Long, H. Zhang, and D. D. Feng, "Fundamentals of content-based image retrieval," *Multimedia Information Retrieval and Management*, pp. 1-26, Springer Berlin Heidelberg, 2003.

- [15] J. R. Smith, and S.-F. Chang, "Transform features for texture classification and discrimination in large image databases," In *Image Processing, Proceedings. ICIP-94., IEEE International Conference*, vol. 3, pp. 407-411, 1994.
- [16] M. N. Do, and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance," *Image Processing, IEEE Transactions*, vol. 11, no. 2, pp. 146-158, 2002.
- [17] W.-Y. Ma, and B. S. Manjunath, "A comparison of wavelet transform features for texture image annotation," *The International Conference on Image Processing*, pp. 256-259, 1995.
- [18] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray scale and rotation invariant texture classification with local binary patterns," In *Computer Vision-ECCV*, pp. 404-420. Springer, 2000.
- [19] Y. An, M. Riaz, and J. Park, "CBIR based on adaptive segmentation of hsv color space," In *Computer Modelling and Simulation (UKSim), 12th International Conference*, pp. 248-251, 2010.
- [20] Z. Dengsheng, and G. Lu, "Evaluation of similarity measurement for image retrieval," *Neural Networks and Signal Processing, Proceedings of the 2003 International Conference*, vol. 2, 2003.
- [21] S. Siegel, "Nonparametric statistics for the behavioral sciences," McGraw-Hill, 1956.
- [22] B. Leo, "Statistics with a view toward applications," Boston: Houghton Mifflin Co., vol. 1, 1973.