# MULTI-TIER CONTENT-BASED MICROSCOPIC IMAGE RETRIEVAL FOR MULTI-IMAGE QUERIES

**J.H.Prasheeba Evangelin**

**Ms.Jemimah simon, M.E**

## Abstract

Image Retrieval is a system of searching, surfing, and retrieving the images from an image dataset. There are two types of different image retrieval techniques namely text based image retrieval and content based image retrieval (CBIR) techniques. Text-Based image retrieval used to manage and retrieve images with traditional database techniques. Content-based image retrieval uses the color feature, shape feature, texture feature, and spatial layout to characterize and manifest the image. This CBIR system is greatly used application in medical filed to handle large set of diseased image database. In this paper, an efficient CBIR system and low-level feature extraction technique is used. This proposed system that gives an efficient retrieving system of microscopic images by multitier approach of CBIR system. Color features and texture features are the features extracted in low-level feature extraction. Normalization is applied over extracted features and for further content-based image retrieval (CBIR) process store the normalized feature. The first tier is classification of disease type with Adaptive neuro fuzzy inference system and the second tier is slide-level image retrieval that deals efficient microscopic image retrieval with k-nearest neighbor (KNN) to find grades of the disease images. The microscopic images taken for this paper are neuroblastoma (NB) and follicular lymphoma (FL) with their subtypes. When using ANFIS instead SVM classifier, the classification of disease images brings more accurateness for retrieval process.

**Keywords**: Image Processing, Feature Extraction, Adaptive neuro fuzzy inference system (ANFIS), Content Based Image Retrieval (CBIR), K-Nearest Neighbor (KNN).

## I.INTRODUCTION

THE medical images provide essential anatomical and functional information about different body parts for detection, diagnosis, treatment planning, and monitoring, as well as medical research and education. Exploration and consolidation of the immense image collections require tools to access structurally different data for research, diagnostics, and teaching. Picture archival and communication systems provide the hardware and software for the storage, retrieval, and management of radiological images such systems use the patient information, and/or modality to index and search the images; the content of the image is not utilized. Content-based image retrieval (CBIR) systems for medical images are important to deliver a stable platform to catalog, search, and retrieve images based on their content. Although several CBIR projects exist for radiology and several other projects are underway there is an acute need for a comprehensive and flexible CBIR system for microscopic images with direct implications for the field of pathology and cancer research.

Microscopic images present novel challenges because they 1) are large in size 2) demonstrate high degree of visual variation due to large variation in preparation (e.g., staining, thickness), and 3) show huge biological variation. Therefore, a well-designed CBIR system for microscopic images can be extremely useful resource for cancer research, diagnosis, prognosis, treatment, and teaching. In other words, such a system can 1) assist pathologists in their diagnosis and prognosis, 2) potentially help to reduce inter- and intrareader variability in clinical practice for the diseases, especially those with complicated classification, 3) help cancer researchers in better understanding of cancer development, treatment monitoring, and clinical trials, and 4) train future generation of researchers by providing consistent, relevant and always available support and assistance.

Neuroblastoma (NB) and follicular lymphoma (FL) tissue images have been collected as part of our project. FL cases are stratified to three histological grades from low- to high-risk category as follows: Grade I, Grade II, and Grade III. NB is the most common extracranial solid cancer in childhood and in infancy.According to International Neuroblastoma Classification System, NB tissues are mainly divided into two subtypes such as stroma rich (SR) or stroma poor (SP) based on the degree of Schwannian stroma development . Additionally, SP tissue has three subtypes such as undifferentiated (UD), poorly differentiated (PD), and differentiating (D). These

subcategories as well as the mitosis karyorrhexis index are used for prognostication. Annotation of microscopic images, e.g., H&E-stained pathology slides, with subtypes of the main disease needs an expert pathologist to select pathology-bearing regions or regions of interests from the whole slide. Then each selected region is annotated semantically by giving a score according to its visual qualitative characteristics. For example, the number of centroblasts ormitotic–karyorrhectic cells can establish a score to interpret the underlying subtype of that disease. The final decision on the grade or subtype of the disease for the whole slide is given after considering the annotations of all sample regions, i.e., the average subtype-related score over all sample regions is assigned as the final score of that whole slide. Considering the extremely large sizes of microscopic images, it is obvious that manual annotation of these images is a time-consuming process and those annotated images may not be easily available for clinical use.

Therefore, one of the aims of this study is to organize the annotated microscopic images in a database and utilize these images for the training of a CBIR system for microscopic images with different disease types and with their subtypes. The novel aspects of our multitiered approach are: 1) it retrieves the most similar disease types in the slide level rather than in the image level by enabling multi-image queries in order to ensure the consistency among the retrieved images, and 2) slide-level scores are weighted in a sophisticated way by modifying the *term frequency*(*tf*)*–inverse document frequency*(*idf*) weighting concepts of information retrieval (IR) theory to decrease the sensitivity of the proposed CBIR system to erroneously annotated sample images in the database. These aspects were designed to mimic the evaluation methodology of pathologists when they review a whole-slide microscopic image. Since in real medical applications, especially for microscopic images at high magnifications, the query object is more likely to be a set of sample images extracted from a whole-slide image rather than being a single image, the multi-image query model suits perfectly for our case. It has been also proved that query by multi-images leads to more scalable and satisfactory query performances by overcoming the limitation on the specification of image content of single-image queries.

In CBIR systems, images are typically represented with feature vectors extracted using low-level image processing techniques. However, similarities in feature vector level does not always guarantee the semantic similarity (i.e., interpretations of images according to their predefined categories) between query image and retrieved images. This is known as the *semantic gap problem*. In this paper, we will explore the effect of slide-level retrieval system with multiple

query images in order to increase the semantic relevance of query image set and retrieved images. It shows the main steps of the CBIR algorithm, e.g., feature extraction, major disease-type classification (first tier), image retrieval according to the subtypes of the diseases (second tier). The rest of the paper is organized as follows.

## II.PROPOSED SYSTEM

The proposed system deals an efficient retrieval of microscopic multi-image queries. The efficiency and accuracy of retrieval system based on feature extraction and classification. The proposed system used multitier content-based image retrieval (CBIR) system for better accuracy. In the dataset, it contains more than one disease tissue microscopic image. The features are extracted from multi image queries by low-level feature extraction method. Then low-level feature extraction extracts the feature by their color feature and texture feature. The hematoxylin and eosin (H&E) images are have limited dominant colors and the color features are extracted which characterize the color and intensity information separately. The texture features are extracted by co-occurrence histogram. The extracted features using RGB, HSV color spaces and intensities are used to calculate co-occurrence histogram in eight directions (i.e., -135, -90, -45, 0, 45, 90, 135 and 180 degrees), to obtain rotation invariant features. The obtained features should be in same magnification level. The extracted color and texture features are concatenated to form feature vector. Then feature vector are normalized. The normalized feature vector is stored for further content-based image retrieval (CBIR) classification. There are two tier used in CBIR system.

In first tier, the disease type is classified by support vector machine (SVM). And in second tier, slide-level image retrieval for microscopic image. The microscopic images taken for this paper are neuroblastoma (NB) and follicular lymphoma (FL) with their subtypes. Compare with existing classify and retrieving system, the proposed system gives better accuracy. There are many factors affecting the performance and accuracy of CBIR systems, such as choosing more discriminative features, similarity measurement criteria, query formulation, and so on. In order to design an effective CBIR system, the initial step in our study is to extract discriminative features from the images in the reference database. These features will also be calculated for query images. One of the most discriminating characteristic of images is color, especially when compared to most common radiological images, which are mostly gray level. Due to the high resolution of microscopic images, subtle changes in characteristics of cells, combinations of

cells, structures, and tissues can also be differentiated from each other by texture characteristics. Therefore, for our CBIR design, we heavily make use of color and texture characteristics and extract these features using low-level image feature extraction techniques.

Our CBIR system operates at two tiers. In the first tier, the designed classifier categorizes the query image/images into one of the major disease types such as FL and NB. Once the disease category of the image is determined, the search for the query image can be carried out among the category relevant subtypes in the subsequent tier. For example, when the query image belongs to NB disease, database images in the first tier will be filtered according to the NB disease category. Then the subsequent search will be only performed on the NB category subset to retrieve the images from the correct category of the query images.

In the second tier, we will use our proposed multi-image query and retrieval methodology to retrieve the images from the reference database in the order of their image-level visual similarities by preserving the slide-level semantic similarity.An SVM-type classifierwas employed to categorize the query image into one of the major disease type such as NB or FL using the extracted features, SVM classifiers are well founded in statistical learning theory and have been successfully used for various classification tasks in computer vision. Their purpose is to find a decision hyperplane for a binary classification problem by maximizing the margin, which is the distance between the hyperplane and the closest data points of each class in the training set that are called support vectors.

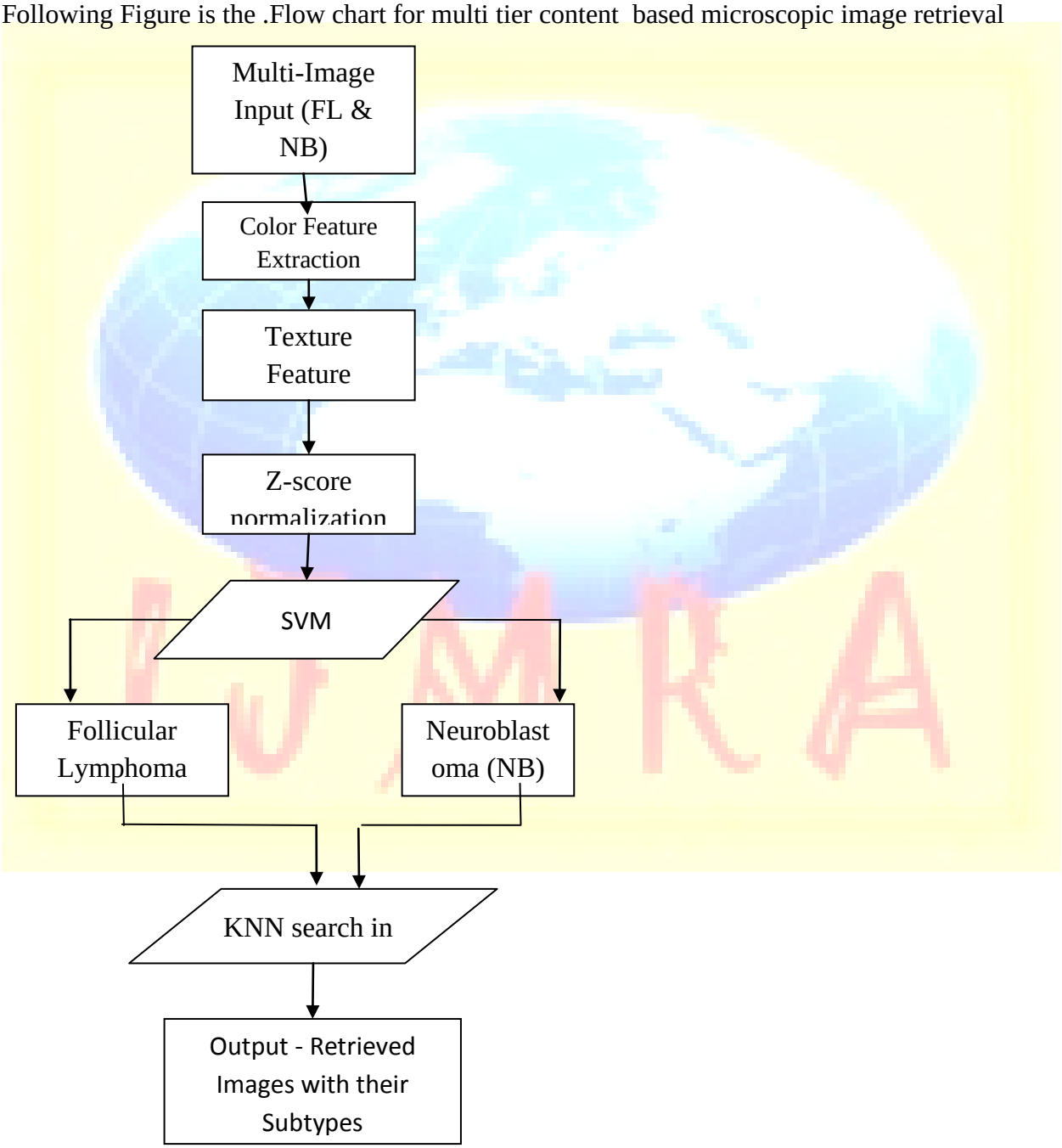
SVM Classifier

The hyperplane is chosen among all the possible hyperplanes through a complex combinatorial problem optimization so that it maximizes the distance (called the margin) between each class and the hyperplane itself. As SVMs are restricted to binary classification, several strategies are developed to adapt them for multiclass classification problems such as one-against-all classification and pairwise classification.

In this part of the CBIR algorithm, we proposed a two-level retrieval system; in the first level, the search is performed similar to traditional CBIR systems such that the images are retrieved based on their image-level similarities. In the second level, the images will be retrieved according to their similarities in the slide level. Once the category of the query image is detected in the first tier, further search is performed on the prefiltered database, which includes only the

sample images of the detected disease category.Each disease has higher level semantic annotations based on their histological grades such as Grade-I, Grade-II, and Grade-III in FL disease or D levels such asSR, UD, PD, and D in NB disease. Therefore, it is necessary to retrieve images related to their higher level semantic characteristics in order to providemore accurate results to the user of the CBIR system.

Following Figure is the .Flow chart for multi tier content based microscopic image retrieval



An SVM model is a representation of the examples as points in space, mapped so that the image samples of the separate categories are divided by a clear gap that is as wide as possible. New image samples are then mapped into that same space and predicted to belong to a category based on which side of the gap they belong. Support Vector Machine (SVM) classifier that classifies the major disease types as Follicular Lymphoma (FL) and Neuroblastoma (NB).

*Cbir System*

Disease specific CBIR systems have been developed for clinical decision support of specific diseases, while some of the CBIR systems were designed for the classification of different types of pathology images, i.e., liver tissue, prostate tissue, breast tissue, lymph node, and so on. Although many promising CBIR approaches were developed for medical applications, there are still gaps in terms of image content, retrieval methodology, performance evaluations, and their application areas. In this paper, the Content based Image Retrieval (CBIR) system is done by two retrieval methods. They are

Image Level Retrieval

Slide Level Retrieval

Each disease has higher level semantic annotations based on their histological grades such as Grade-I, Grade-II, and Grade-III in FL disease or D levels such as SR, UD, PD, and D in NB disease. Therefore, it is necessary to retrieve images related to their higher level semantic characteristics in order to provide more accurate results to the user of the CBIR system.

*Image Level Retrieval*

The Image level retrieval algorithm, first initialize the scores of the query image as zero and then find the correlation distance measure between each and every query images set with all image databases.

$$Dist\left(F_{Q^n}, F_{DS^t}\right) = 1 - \{Correlation\{F_{Q^n}, F_{DS^t}\}\}$$

$$Dist(F_{Q^n}, F_{DS^t}) = 1 - \left\{ \frac{\langle (F_{Q^n}), (F_{DS^t}) \rangle}{\|F_{Q^n}\| \|F_{DS^t}\|} \right\}$$

$$t = 1, \ldots, T, T = |F_{DS}| \; and \; n = 1, \ldots, N, N = |Q|$$

Where N is the number of individual query entities in the given query image set Q, T is the number of images in the reference dataset DS, and $F_{Q^n}$ represents the feature vector of the *n*th query image, $F_{DS^t}$ represents the feature vector of the *t*th image of the given dataset, $\langle .,. \rangle$ is the inner product, $\| . \|$ is the $L_2$ norm, and $| \cdot |$ is the cardinality. This algorithm updates the scores to most similar images to the database images, the scores are summed the number of occurrences of each image in the dataset for a k-nearest neighbor (KNN) search of that query image set. This image level retrieval algorithm results finding most similar images from the query images and image database.

*Slide Level Retrieval*

In slide level retrieval algorithm, introduce the conventional way of ranking the similarity of slides to a given query image set is by sorting the similarity scores of the reference slides independent from their subtypes and retrieving the highest scored slides from the database, which means that subtypes of the slides are considered equally important.

The scores are sorted in descending order and using the score, the rank of the images calculated with their subtype number. At last, the weighted score is calculated by multiplying corresponding score of image, subtype of the image and rank weight of the image and then it should be sorted in descending order. And displays the retrieved image subtypes using the highest scored slides. This Content based Image Retrieval (CBIR) system retrieves most similar subtype

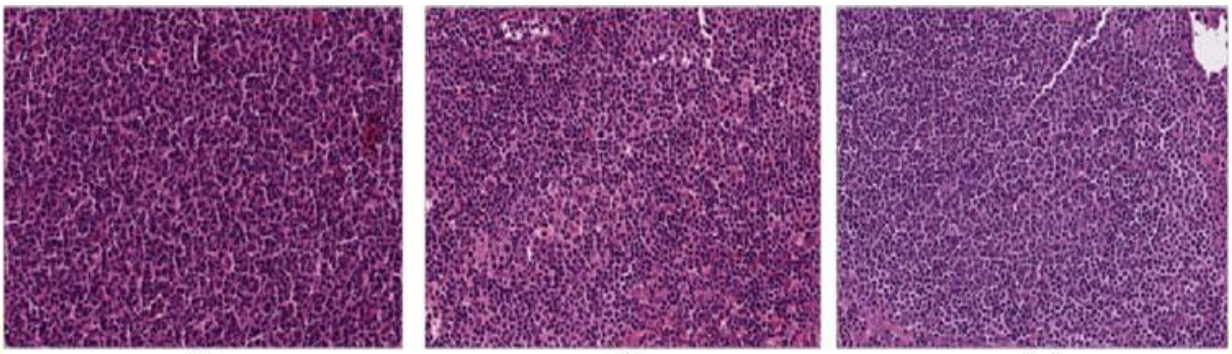



Fig Sample H&E-stained FL images associated with thethree grades. (a) Grade I. (b) Grade II. (c) Grade III.

.

## III CONCLUSION:

In this paper, we have presented a novel content-based microscopic image/slide retrieval algorithm. We have demonstrated that by using the proposed weighting scheme inspired by IR theory, the slide-level retrieval performance of the CBIR system is considerably better than the traditional image-level retrieval accuracy for all seven subtypes of two challenging diseases, which have inter- and intrareading semantic variations, intraslide semantic variations, and intersubtype visual similarities. This CBIR system can enable the user, e.g., a pathologist, to select multiple HPF regions from a suspected tissue and submit those images as a query to the CBIR system and retrieve the most relevant slides with their semantic annotations with higher accuracies. The results, achieved under those challenging conditions, are also promising for automatic and unsupervised selected query images based on their HPF regions. Application of the proposed weighting strategy, inspired by the IR theory, is not limited to microscopic images only, and can be also useful for any type of multiquery search and content-based retrieval systems.

REFERENCES

[1] R. H. Choplin, J. M. Boehme, and C. D. Maynard, "Picture archiving and communication systems: An overview," *Radiographics*, vol. 12, no. 1, pp. 127–129, 1992.

[2] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications clinical benefits and future directions," *Int. J. Med. Informat.*, vol. 73, no. 1,pp. 1–23, 2004.

[3] W. Hsu, L. R. Long, and S. Antani, "Spirs: A framework for content-based image retrieval from large biomedical databases," *Stud. Health Technol. Informat.*, vol. 129, no. Pt. 1, pp. 188–192, 2007.

[4] H. L. Tang, R. Hanka, and H. H. S. Ip, "Histological image retrieval based on semantic content analysis," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 1, pp. 26–36, Mar. 2003.

[5] C.-R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick, "Assert: A physician-in-the-loop content-based retrieval system for HRCT image databases," *Comput. Vis. Image Understand.*, vol. 75, no. 1–2, pp. 111–132, 1999.

[6] L. Zheng, A. Wetzel, J. Gilbertson, and M. Becich, "Design and analysis of a content-based pathology image retrieval system," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 4, pp. 249–255, Dec. 2003.

[7] L. Yang, O. Tuzel,W. Chen, P.Meer, G. Salaru, L. Goodell, and D. Foran, "Pathminer:Aweb-based tool for computer-assisted diagnostics in pathology," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 3, pp. 291–299, May2009.

[8] S. A. Napel, C. F. Beaulieu, C. Rodriguez, J. Cui, J. Xu, A. Gupta, D. Korenblum,H. Greenspan, Y. Ma, and D. L. Rubin, "Automated retrieval oct images of liver lesions on the basis of image similarity: Method and preliminary results.,"
*Radiology*, vol. 256, no. 1, pp. 243–252, 2010.

## BIBLIOGRAPHY

**Mrs.Prasheeba Evangelin.,** received her B.E degree from Vins Christian College Of Engineering, Chunkankadai, Tamil Nadu, India and is currently pursuing her M.E degree in Vins Christian College Of Engineering , Chunkankadai , Tamil Nadu ,India.

Her area of interests are Image processing, Data Mining and Networks. She has attended International Conference at India and at Bharathiyar Institute Of Technology For Women, Salem, India.

**Ms. Jemimah Simon. , M.E** received her B.E degree from St.Xavier's Catholic College of Engineering , Tamil Nadu ,India and her M.E from Adhiyamaan College of Engineering, Hosur, India

Her area of interests are Image processing, Data Mining and Networks. She is currently working as an Assistant Professor in Vins Christian College Of Engineering , Chunkankadai ,Tamil Nadu, India.