International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

BINARY DESCRIPTOR BASED MULTI-CLASS OBJECT CATEGORIZATION FOR THYROID NODULE IMAGE ANALYSIS

T.Manivannan[1], Dr.A.Nagarajan[2*], Dr.S.Santhoshkumar[3]

[1]Research Scholar, [2&3]Assistant Professor*

[1&2] Department of Computer Applications*

[3]Department of Computer Science

[1,2 &3]Alagappa University- Karaikudi.*

Abstract

In this research paper, an improved method to identify the object class in thyroid nodule visual images using multi-class classification is presented. A general method for object categorization is to quantize feature descriptors into visual words and then form a bag-of-feature model. Bag of Visual Word (BoVW) based Image Retrieval is a well-received method for Content Based Image Retrieval (CBIR). After this, multi-class classification is used to categorize objects into different classes. Feature detection and description act as the backbone of multi-class classification. BRISK & FREAK (Binary Robust Invariant Scalable Key points and Fast Retina Key point) are well-known binary descriptors in many computer vision applications. in this research paper proposed to use the combination BRISK with BRISK (BRISK-BRISK) and BRISK with FREAK (BRISK-FREAK) as binary detector and descriptor under the framework of SVM (Support Vector Machine) classification. The classification of thyroid nodule image visual categories using the proposed algorithm and their performance is evaluated based on the confusion matrix. A comparison of these methods with different kernels of the SVM classifier is also observed.

Keywords - *Bag-of-words, Bag of Visual Word (BoVW), Binary descriptors, Image classification, SVM (Support Vector Machine) Object recognition*

1. Introduction

Booming technologies like Internet, Digital and social media make the availability of multimedia data in plenty. In particular, accessing image data has become an essential thing and thereby many applications like Object Recognition, Object Detection, and Image Retrieval need to be performed efficiently and accurately. Currently, BoVW based Image Retrieval [17] is one of a successful way for retrieving images. BoVW represents an image as collection of frequency occurrences of visual words. A basic technique used for the implementation of image categorization is a bag of words (BoW) [1]. A visual multi-class categorization using bag-of-keypoints originated from the world of document retrieval [2]. The elementary difference between BoW for text and image classification is that visual words are generated artificially after feature vector quantization [3], whereas text words are formed through natural sampling in proportion to language context [4]. The Bag-of-word in image

Received: August 06, 2025

classification is therefore known as Bag-of-visual-word (BoVW) and the process generates a histogram of the occurrence of a visual word in an image.

Apart from its ease of use it comes with shortcomings like lack of Spatial Information, Low Discriminative Power, and Quantization problems etc. As BoVW holds only frequency occurrence of visual words, it does not provide any spatial information and thereby it declines the discriminative power of identifying correct images. Quantization Error occurs in feature quantization process while mapping a feature to a single nearest visual word. It causes polysemy which means same visual word is tagged to semantically different object and synonymy is caused by assigning different visual words for semantically similar objects. Polysemy happens when the vocabulary size is small and synonmy occurs for large vocabulary size and thus responsible to reduce the precision and recall values.

2. Related Works

To include more spatial information using BoVW based image retrieval the authors [23] inspected the co-occurrence of Geometry Preserving Visual Phrase (GVP) of visual words. Along with this they increased the retrieval accuracy by applying Minhash technique for image indexing and experimented to show better performance when compared to native BoVW and RANSAC. With a view to indulge spatial contextual information the authors [12] proposed a new representation and enforced soft quantization for query image. Their main idea was that the visual words belonging to the same objects are placed in the neighboring bins where the distracter words are located in further away bins. With this motto, visual phraselets which contains group of visual phrases was introduced for measuring the similarity among the images. It does filter of false matches and thereby improved the discriminative power for large scale image retrieval to 54.6% over the base line approach.

The whole process is vitally trifurcated as the following parts,

i) Feature extraction and description

ii) Bag of word

iii) Classification

2.1 Feature Extraction and Description

The favored element of an image is referred to as a feature. Feature extracted have to be invariant to scale and rotation in an effort to enhance the class performance. Feature extraction is broadly classified into two parts: Global feature and Local feature. The global feature focuses on the entire image and considers all pixels in an image, whereas local features use interest points set in the image and extract the image properties around these feature points [5].

The various local feature descriptors can be broadly classified as gradient-based and binary descriptors. HoG (Histogram of Oriented Gradient) [6] and SIFT (Scale Invariant Feature Transform) [7] are well-known gradient-based feature detectors which are scale, rotational and illumination invariant features. SURF (Speed Up Robust Features) [8], another powerful descriptor is the fast implementation of SIFT using integral images.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Gradient-based approaches are based on the histogram of a gradient in which a patch gradient is calculated for each pixel, which is computationally expensive. These descriptors impose a large computational burden for real-time systems. In order to overcome this, binary descriptors can be used for real-time applications. In binary descriptor, a binary string is formed by comparison of intensity values that provide information about image patches. High speed in computation can be achieved can be done very fast as only the comparison between image intensities needs to be computed. Binary descriptors are constructed of three things: A sampling pattern, orientation compensation and sampling pairs. BRISK [9] and FREAK [10] are most the popular binary descriptors.

2.2 Bag of word

Bag of Word Method originally obtained from NLP [11]. The bag of keypoints the way of dealing with visual categorization. The vector quantization, as the result of visual words, is prepared. It simply represents the histogram that is intentional on the basis of the variety of particular image pattern occurrences.

2.3 Classification

The feature vector obtained by bag-of-keypoint is fed into a classifier, which decides the assignment of the image to the one out of the remaining classes or categories, to which it belongs. Such a type of classifier is known as a multi-class classifier. Naive Bayes Classifier and Support Vector Machine [11] have been used for classification and show Support Vector Machine (SVM) is simple and works better than the Naive Bayes Classifier. SVM is used to classify images based on its Bag-of-visual-Words.

SVM with a linear kernel can be represented as:

$$K(x,y) = xy + c \qquad (1)$$

where x and y are vectors in the input space, i.e. features vector computed from training samples. SVM classifier with the linear kernel is simple, fast and computationally efficient for training as well as for classification purposes. SVM with nonlinear kernels such as Gaussian kernel and polynomial kernel gives improved performance.

For two samples x and x', represented as feature vectors in some input space, the Gaussian kernel can be defined as,

$$K(x, x') = \exp(-\Upsilon \| x - x' \| 2) \qquad (2)$$

Where $\Upsilon = 1/22\sigma$ in which σ is a free parameter. The polynomial kernel for degree d is given by:

$$K(x, x') = (xy + c)d \qquad (3)$$

Where d is the degree of the polynomial.

For multi-class categorization, SVM is used as a multi-class classifier [12], which works on the principle of one-versus-all, for example, if four different classes are present, then it will be considered as four sets of binary classification problems i.e. class 1 will act as positive class and the other three classes together form the negative class. The same process is repeated for each of the other classes to form four separate 2 class classification problems.

3. Proposed Methodology

The general image categorization system's block diagram is shown in figure 1. A collection of thyroid nodule images are considered as training data for Keypoint detection and description. A group of predetermined clusters using K-mean clustering with the help of vocabularies from feature descriptors may have the vector quantization due to the quantization of the feature vector.

The outcome of the clustering, the resulting centers would be taken as visual words. The visual word frequency curves help create a histogram that shows the contents of an image. Therefore, a histogram of visual words is generated from each training image, so an image could be displayed as a bag of visual words. The multi-class classifier is trained using the feature vector generated by the bag-of-keypoints [5], which are trained to decide the assignment of an image to an element category.

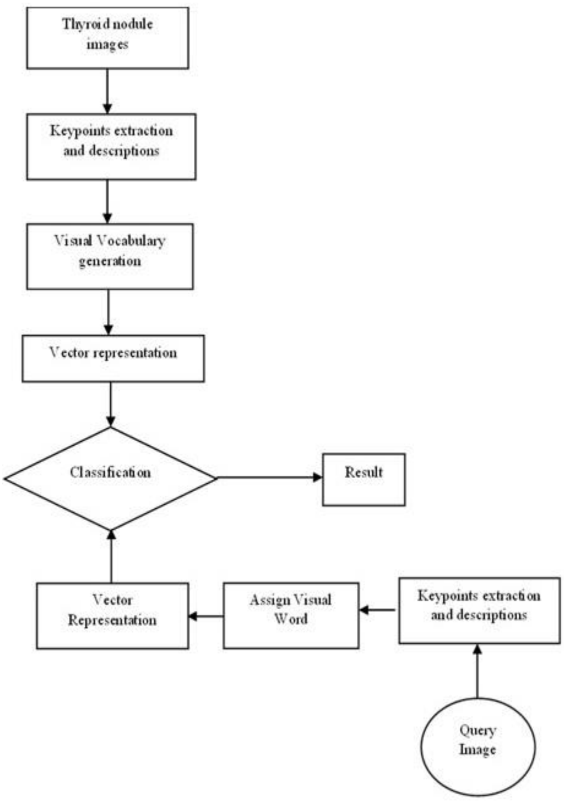

Fig. 1. Block diagram for multi-class classification system

During the test phase, feature points are extracted from a test image and the feature descriptor is calculated for each feature point. The test image is presented as a histogram of visual words. The classifier categorizes the check image based totally on this histogram of visual words. The vocabulary size should be adequate and sufficient to distinguish among changes in images. however, it needs to not be too large as it may differentiate irrelevant changes or records in images such as noise.

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The classification of the image analyzes the numerical properties of different image functions and organizes data into categories. The paper proposed a method of image categorization of the visual objects with the help of binary descriptors: BRISK and FREAK. The method is performed on the subset of the TI-RADS - Thyroid Imaging Reporting and Data System database [13]. 3 thyroid nodule scan object classes are considered. There is no restriction on choosing any particular number of categories or any particular object category. One can randomly choose the number of object classes. 40 percentages of randomly chosen images are used for training and the remaining 60 percent are used for validation from each category.

The feature detection of training images is done using BRISK. Fig. 2, 2.1, and 2.2 shows an example of a plot of BRISK detection points after considering 100,300,500 strongest points. The descriptions of feature points obtained are performed using BRISK and FREAK binary descriptors separately. The extracted feature descriptors from training images of each category are being used to form a bag of features or visual vocabulary. The K-mean clustering algorithm is used on the feature descriptors extracted. It is an iterative algorithm and it repeatedly groups the descriptors into a pre-defined number of clusters (K=500). The resulting clusters formed are compact, mutually exclusive and separated by similar characteristics. Each cluster center is considered as the feature or visual word and hence forms a visual vocabulary.

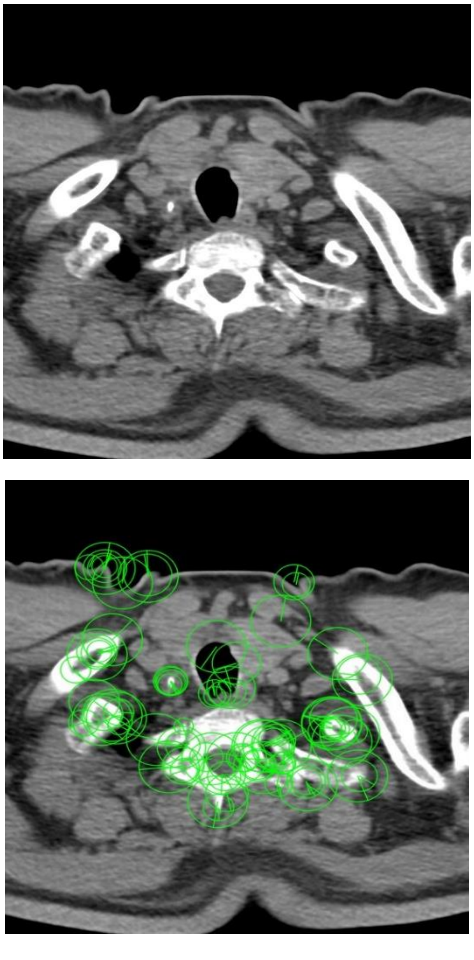

Fig. 2. Thyroid nodule Image with BRISK feature point (strongest 100 points)

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

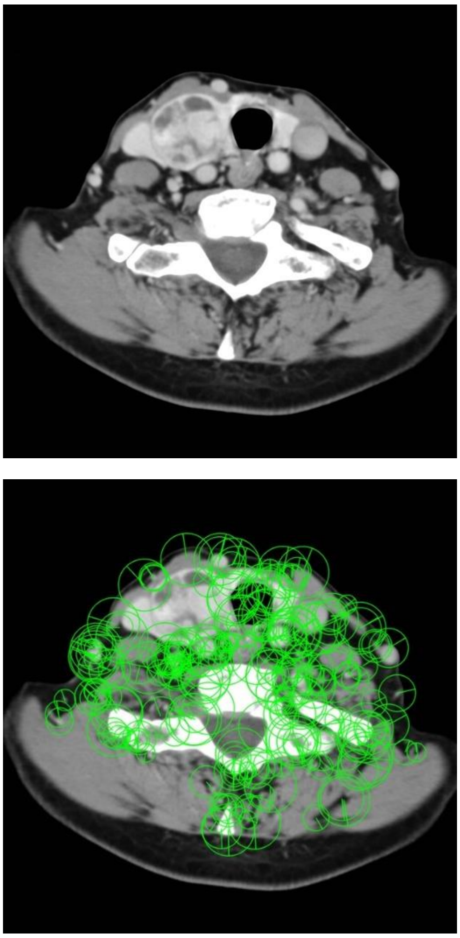

Fig. 3. Thyroid nodule Image with BRISK feature point (strongest 300 points)

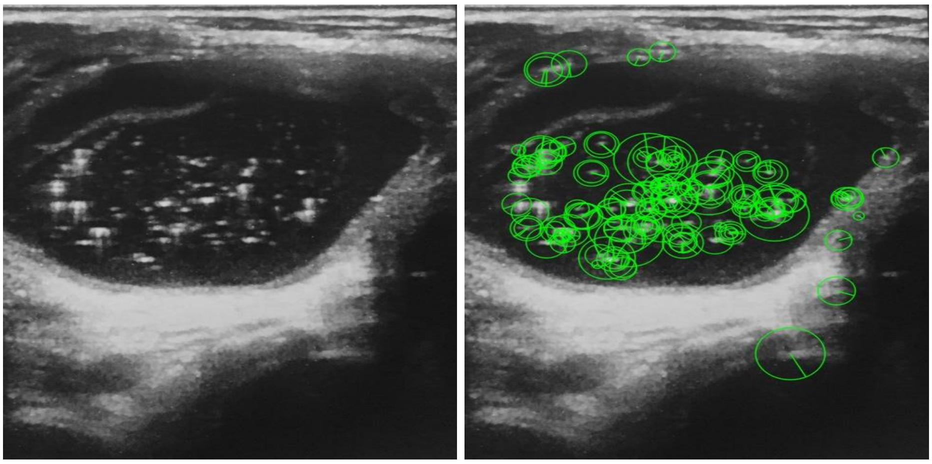

Fig. 4. Thyroid nodule Image with BRISK feature point (strongest 500 points)

Each image of the training set is then encoded using this bag of features and a histogram is constructed for each of the images. Encoding is being done using an approximate nearest neighbor algorithm after extracting features and incrementing the histogram bins for each input training image. The vocabulary size will decide the size of the histogram (500 in our case) hence; the histogram size equals the number of visual words or the number of cluster centers. Thus, each image is now to be the result of a histogram and this will be act as a

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

feature vector which is used as training data. Fig.3 shows the example of images and their histogram representation.

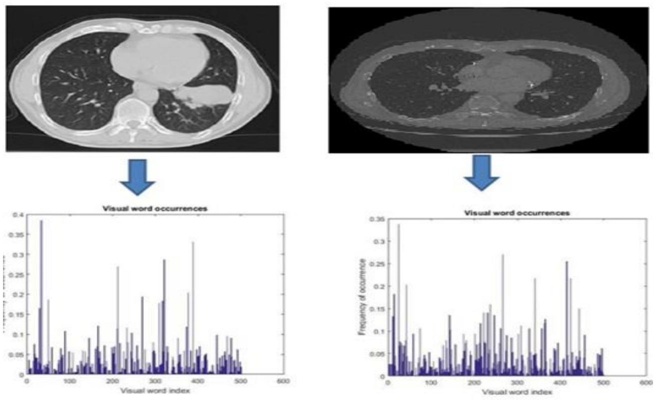

Fig. 5. Visual word representation of the image with histogram construction

Once the descriptor has been assigned to form the feature vectors, we reduce the problem of preferred visual categorization to that of multi-class supervised learning, with as many classes as there are visual categories defined [14]. To train the classifier using a bag of the word, errors-correcting output codes (ECOC) structure with binary support vector machine (SVM) classifiers is used to train a multi-class classifier which became introduced [15]. This is only a coding strategy, once in a while called one-vs-all or one-vs-rest. The excellent of a classifier is evaluated by means of the usage of the classifier against the validation image set and a confusion matrix is formed, that represents prediction analysis.

4. Performance Evaluations

The performance is evaluated on the basis of a contingency table (confusion matrix), where diagonal elements represent the TP (True Positive) class i.e. the object is classified to the correct category to which it belongs and off-diagonal elements are the errors i.e the classifier categorizes the object to the wrong class [16]. As four categories are taken, the confusion matrix must be 4*4 size. The performance evaluation in image categorization depends upon parameters such as accuracy, precision, recall and F-score.

4.1 Accuracy, Precision and Recall

Accuracy [17] is the proportion of the overall number of predictions that were correct. It’s far represented using the following equation:

$$\text{Accuracy} = TP + TN / TP + TN + FP + FN \qquad (4)$$

Precision [18] is a high-quality predictive value is the proportion of decided on magnificence. It is proven in equation (5). Where TP = true positive, FP = false positive.

$$\text{Precision} = TP / TP + FP \qquad (5)$$

Recall corresponds [19] to the authentic positive rate or sensitivity measures. it's far the proportion of positives which are efficaciously identified may be calculated by using the use of equation:

Received: August 06, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Recall = TP / TP + FN (6)

4.2 F-score

The combination of precision and recall leads to a new measure known as F-score [20]. It combines precision and recall by taking the harmonic mean of both and can be represented as the equation (7).

Fscore = 2 * Precision:Recall / Precision + Recall (7)

Accuracy and F-score are used for evaluating the performance of the proposed method.

5. Results and Discussions

The proposed methodology is tested on a publically out there on the subset of the TI-RADS - Thyroid Imaging Reporting and Data System database. The software used is MATLAB 2024b with 8 GB RAM. We have compared the performance of two binary descriptors BRISK and FREAK, which are used for a bag of visual word-formation. Two combinations of keypoint detector descriptors using BRISK-BRISK (BB) and BRISK-FREAK (BF) are used and compared for image categorization into the four classes. Performance evaluation is done on the basis of confusion matrix formation of the whole system using different classification kernels. Table 1 & 2 shows the confusion matrix for linear kernel for BB and BF respectively.

Table 3 & 4 shows the confusion matrix for the Gaussian kernel for both the descriptors. The confusion matrix for BB and BF using a polynomial kernel is shown in Tables 5 & 6.

Table 1. Linear (BRISK-BRISK)

True Classes >	Image 1	Image 2	Image 3
Thyroid Image 1	0.92	0.00	0.01
Thyroid Image 2	0.03	0.97	0.00
Thyroid Image 3	0.04	0.03	0.93

Table 2. Linear (BRISK-FREAK)

True Classes >	Image 1	Image 2	Image 3
Thyroid Image 1	0.99	0.00	0.00

Received: August 06, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Thyroid Image 2	0.05	0.85	0.08
Thyroid Image 3	0.04	0.01	0.93

Table 3. Gaussian Kernel (BRISK-BRISK)

True Classes >	Image 1	Image 2	Image 3
Thyroid Image 1	0.93	0.00	0.01
Thyroid Image 2	0.00	0.97	0.01
Thyroid Image 3	0.01	0.07	0.88

Table 4. Gaussian Kernel (BRISK-FREAK)

True Classes >	Image 1	Image 2	Image 3
Thyroid Image 1	0.95	0.00	0.00
Thyroid Image 2	0.04	0.96	0.00
Thyroid Image 3	0.03	0.04	0.93

Table 5. Polynomial Kernel (BRISK-BRISK)

True Classes >	Image 1	Image 2	Image 3
Thyroid Image 1	0.91	0.00	0.03
Thyroid Image 2	0.01	0.97	0.01
Thyroid Image 3	0.01	0.04	0.95

Received: August 06, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 6. Polynomial Kernel (Brisk-Freak)

True Classes >	Image 1	Image 2	Image 3
Thyroid Image 1	0.97	0.00	0.01
Thyroid Image 2	0.04	0.86	0.04
Thyroid Image 3	0.01	0.01	0.96

The confusion matrices obtained show that the polynomial kernel gives a great result in terms of classification accuracy while in comparison to linear and Gaussian kernels, for both BB and BF methods, as shown in Table 7.

Table 7. Comparison of Accuracy With SVM Classification Kernels

SVM Classification Kernels	BRISK-BRISK (Accuracy %)	BRISK-FREAK (Accuracy %)
Linear	84	85
Gaussian	85	86
Polynomial	91	93

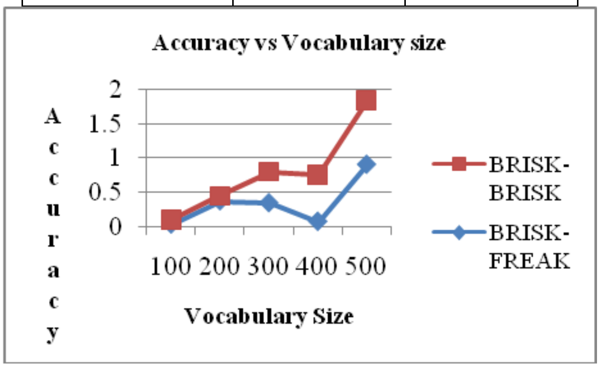

Fig.6. Accuracy versus Vocabulary Size graph of BRISK-BRISK and BRISK-FREAK for polynomial kernel

The impact of vocabulary size on accuracy is shown in Fig.6 for both cases i.e BRISK-BRISK and BRISK-FREAK. It shows that the increment in vocabulary size causes an increment in accuracy in both cases. For vocabulary size 500, we obtained the highest accuracy is 91 percent and 93 percent for BB and BF respectively. It is observed that BF outperforms BB in all classification kernels, however, a non-linear kernel classifier shows better accuracy.

Fig.7 shows the impact of changing vocabulary size on the F-score. As the vocabulary size increases, the F-score also increases. BB gives F-score of 0.9 and BF gives 0.92 which shows that BF is having better performance than BB.

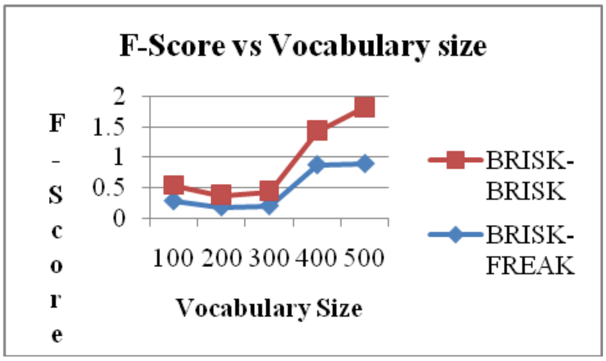

Fig.7. F-score versus Vocabulary Size graph of BRISK-BRISK and BRISKFREAK for polynomial kernel

6. Conclusion

In this research paper, an improved method is proposed for image classification and how binary descriptors can be used for visual image categorization using the Bag-of-feature method. This method has been evaluated on three categories of the TI-RADS - Thyroid Imaging Reporting and Data System database. Our results show that BRISK and FREAK feature descriptors are simple, fast and have a low computational complexity than gradient-based descriptors. We have taken BRISK as the interest point detector. With the interest points obtained, both BRISK and FREAK are used as descriptors separately. These two descriptors were compared with different classification kernels: linear, gaussian and polynomial. FREAK descriptor worked better than BRISK descriptor in all three cases. However, there is a small trade-off between accuracy and computational complexity. Polynomial kernel showed more accuracy and F-score as compared to linear and gaussian kernels. But it is more complex than the other two kernels.

References

[1] Raghavendra et al., "Optimized multi-level elongated quinary patterns for the assessment of thyroid nodules in ultrasound images". *Comput. Biol. Med.* Volume 95, 2018, pp. 55-62.

[2] Cheng Xie et al., "Ultrasonography of thyroid nodules: a pictorial review", *Insights Imaging,* 7(1): doi: 10.1007/s13244-015-0446-5, 2016, pp. 77–86.

[3] P. Civicioglu, "Backtracking search optimization algorithm for numerical optimization problems", *Applied Mathematics and Computation,* volume 229, 2013, pp. 8121 – 8144.

[4] M. Varma and A. Zisserman. (2002). Classifying images of materials: Achieving viewpoint and illumination independence. *Computer VisionECCV 2002*. Springer, pp. 255–271.

[5] G. Csurka et al. (2004). Visual categorization with bags of keypoints. *Workshop on statistical learning in computer vision, ECCV*, vol. 1, no. 1-22. Prague, pp. 1–2.

[6] Alahi et al. (2012). Freak: Fast retina keypoint. *Computer Vision and Pattern Recognition (CVPR*), pp. 510–517.

[7] C. Barata et al. (2014). Two systems for the detection of melanomas in dermoscopy images using texture and color features. IEEE Systems Journal, vol. 8, no. 3, pp. 965–979.

[8] F. Riaz et al. (2017). Content-adaptive region-based color texture descriptors for medical images. *IEEE Journal of Biomedical and Health Informatics,* vol. 21, no. 1, pp. 162–171.

[9] Q. Li et al. (2014). Medical image classification with convolutional neural network. Proceedings of 13th International Conference on *Control Automation Robotics & Vision (ICARCV),* IEEE, Singapore, pp. 844-848.

[10] Y. Guo et al. (2016). Dynamic network surgery for efficient DNNs. *Advances in Neural Information Processing Systems,* MIT Press, Cambridge, MA, USA, pp. 15-24.

[11] G. Ma et al. (2016). Multi-feature fusion deep networks. *Neurocomputing,* vol. 218, pp. 164–171.

[12] M. Gong et al., "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering", *IEEE Trans. Image Process.*, volume 21, no. 4, 2012, pp. 2141–2151.

[13] Anam Mustaqeem et al., "An Efficient Brain Tumor Detection Algorithm Using Watershed & Thresholding Based Segmentation", *Image Graphics and signal Processing,* 2012, pp. 34-39.

[14] C. Li et al., "A level set method for image segmentation in the presence of intensity in homogeneities with application to MRI", *IEEE Trans. Image Process*, volume 20, no. 7, 2011, pp. 2007–2016.

[15] S.Rahnamayan and Z.S. Mohamad, "Tissue Segmentation in Medical Images Based on Image Processing Chain Optimization", *IEEE International Workshop on Real Time Measurement, Instrumentation & Control,* 2010.

[16] D. Karaboga and B. Basturk, "An artificial bee colony (abc) algorithm for numeric function optimization", *IEEE Swarm Intelligence Symposium, Indianapolis, Indiana, USA, 2006.*

[17] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," *Computer Engineering Department, Erciyes University, Turkey, Tech.* Rep. TR06, 2005.

[18] H.-C. Shin et al. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298.

[19] M.-E. Nilsback and A. Zisserman. (2006). A visual vocabulary for flower classification. *Computer Vision and Pattern Recognition IEEE Conference,* pp. 1447–1454.

[20] J. Philbin et al. (2007). Object retrieval with large vocabularies and fast spatial matching. *Computer Vision and Pattern Recognition IEEE Conference,* pp. 1–8.

Received: August 06, 2025