Hato et al. *Iraqi Journal of Science, 2024, Vol. 65, No. 8, pp: 4606-4622*
DOI: 10.24996/ijs.2024.65.8.38

ISSN: 0067-2904

Comparative Analysis for Bag of Features (BoF) Performance

Eman Hato[1], Zinah S. Abduljabbar [1], Zainab J. Ahmed [2]*
[1]*Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq*
[2]*Department of Biology, College of Science, University of Baghdad, Baghdad, Iraq*

Received: 27/2/2023 Accepted: 1/8/2023 Published: 30/8/2024

Abstract

The accuracy of the Bag of Features (BoF) is greatly affected by the discriminatory power of feature extraction techniques. This paper presented a proposed method designed to find the best features technique for constructing a BoF model according to the image classification. It consists of four stages: feature extraction, where detectors and descriptor feature techniques have been exploited to generate different BoF models. Each BoF model is generated depending on what detector and descriptor are used. The BoF models are constructed to represent the images as feature vectors. The classification process is then performed on two image datasets. Finally, the efficiency of BoF models is analyzed and evaluated with respect to the accuracy of their classification performance. Experimental results indicated that the best level of accuracy was provided by the proposed BoF model with the KAZE features method. The results also showed that the BoF model with Speeded Up Robust Features (SURF) was superior to other feature methods in terms of execution time, which was 0.01218 seconds. Moreover, the BoF model generated by the SURF detector combined with the KAZE descriptor achieved a high level of accuracy of 0.99 and kept the time complexity low (0.01948 seconds).

Keywords: BoF, Features Extraction, KAZE, Sensitivity, Specificity.

التحليل المقارن لأداء حقيبة الميزات

ايمان هاتو[1]*, زينه صادق عبد الجبار[1], زينب جواد احمد[2]
[1]قسم علوم الحاسوب ، كلية العلوم ،الجامعة المستنصرية، بغداد، العراق
[2] قسم علوم الحياة ، كلية العلوم ،جامعة بغداد، بغداد، العراق.

الخلاصة

تتأثر دقة حقيبة الميزات بشكل كبير بتقنيات استخراج الميزات المستعملة. في هذه الورقة تم اقتراح طريقة صممت من أجل إيجاد أفضل تقنية استخراج الميزات لبناء نموذج حقيبة الميزات بناءً على عمليات تصنيف الصور. تتكون الطريقة المقترحة من أربع مراحل وهي: استخراج الميزات حيث تم استعمال تقنيات الوصف والكشف لإنشاء نماذج مختلفة من حقيبة الميزات. كل نموذج يتم توليده بناءً على اي من تقنية الكاشف و الواصف سيستعمل. ان هذه النماذج تمثل مصفوفات الميزات التي استعملت لتمثيل الصور. بعد ذلك يتم تنفيذ عملية التصنيف على مجموعتين من البيانات الصورية. وفي المرحلة الاخيرة، يتم تحليل وتقييم كفاءة نماذج حقيبة الميزات المقترحة بالاعتماد على دقة أداء التصنيف. أشارت النتائج التجريبية إلى أن نموذج حقيبة

*Email: zainab.jawad@sc.uobaghdad.edu.iq

الميزات المصمم باستعمال واصفات KAZE أظهر أفضل قيمة لدقة التصنيف. كما بينت النتائج ان نموذج حقيبة الميزات باستعمال الكاشف SURF كان متفوقًا على طرق الميزات الأخرى المستعملة من حيث وقت التنفيذ والذي كان 0.01218 ثانية. علاوة على ذلك ، أن نموذج حقيبة الميزات الذي تم إنشاؤه بواسطة كاشف SURF مع واصف KAZE حقق مستوى عالٍ من الدقة والتي كانت قيمتها 0.99 مع الحفاظ على التعقيد الزمني منخفضًا (0.01948 ثانية).

1. Introduction

Classification of scenes is a fundamental process of human vision that allows us to efficiently and rapidly analyze our surroundings. Image classification is the process of assigning the category the image falls under. This topic has attracted increasing interest as a key component in many applications of computer vision, for example, image classification, matching, object recognition, object tracking, human action recognition, and image and video retrieval [1].

Image classification is considered a very difficult task for computer programs (or machines). Major difficulties include complex and hard-to-describe objects in an image, objects occluding other objects, and the gap between arrays of numbers representing physical images and conceptual information perceived by humans [2].

The most important steps in various image classifications include the determination of suitable classifiers and feature extraction. The features must be extracted carefully from an image to obtain the most relevant information about the image content, find robust descriptors, and preserve the representation of the entire image. For example, the shape may be a good feature to distinguish between boats and cars, but it is not good to distinguish between coast and forest [3, 4]. Many image classification approaches are based on local features to obtain better details of the image. Local features detect interest points in an image and describe small neighborhoods around them using a set of vectors. Different approaches have been proposed to detect and describe the interest points, for example, Scale Invariant Feature Transform (SIFT), Histograms of Oriented Gradient (HoG), Binary Robust Invariant Scalable Keypoints (BRISK), and SURF [5, 6].

The number of local features for each image is huge. To overcome this drawback, the extracted features are encoded in the image representation vectors by feature coding methods such as Bag of Features (BoF), Fisher Vector (FV), or Locality Constrained Linear Coding (LLC) [7]. The BoF is a dictionary-based method to describe an image that can provide a higher-level representation. The histogram is a result of the BOF procedure for feature extraction that reflects feature distribution, which leads to describing the image properties [8].

2. Bag of Features (BoF)

BoF (also known as “Bag of Visual Words”) is a method to represent the features of images. BoF is derived from the Bag of Words (BoW) model that was used in document classification and retrieval, where the occurrences of words are used as vector features. In computer vision, BoF can be used for image classification and retrieval by treating image features as words. Thus, BoF is a vector of occurrence numbers of a vocabulary of local image features [9].

The BoF has become popular in recent years thanks to its effectiveness and the quality of its results. The BoF model comprises three main stages, including feature extraction, codeword generation, and feature coding. The procedure for creating the BoF model can be summarized as follows: Firstly, the points or regions of interest are detected. Then, for each detected point or region, a fixed feature vector representation for image content around the detected keypoint is built (i.e., a descriptor). After that, the descriptors are quantized into a predetermined visual vocabulary (also known as the codewords or codebook) using the clustering technique. Lastly, the occurrences for each specific visual vocabulary are computed to construct the BoF model, namely the histogram of visual word frequencies. Figure 1 illustrates the general structure of BoF [10, 11].

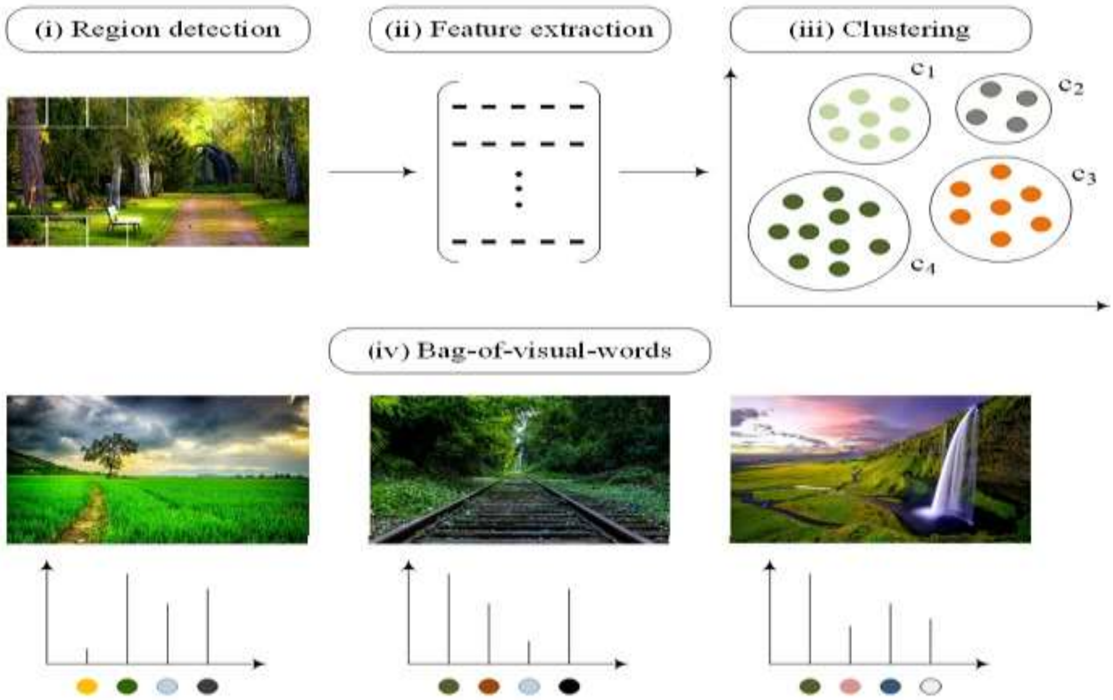

Figure 1: Bag of Features Model **[11]**

Despite its good performance, the BoF suffers from some limitations. The loss of spatial information during the feature extraction and feature coding stages is a crucial limitation of BoF. In other words, the order of the features in the image is lost during the feature extraction process [12]. This led to the proposal of many improvements and developments in every step of BoF modeling.

3. Related Work

A comprehensive survey of existing BoF techniques and their application areas is presented as follows:

D. Srivastava et al. [13] proposed a method for image classification using BoF. The SURF is used to detect interest points that are used to extract regions of interest (ROI). The features of LBP are computed for ROIs that are later clustered using Clustering with Fixed Centers (CFC) to generate BoF vectors. The average accuracy achieved with the SVM classifier is 81.7.

R. Mandal et al. [14] presented a method for document retrieval using handwritten signature recognition. The features are extracted based on the BoF that used the SIFT descriptor. Then, the support vector machine (SVM) classifier is used to recognize the signature components in the document. Lastly, the signature components are grouped and matched with the query signature to retrieve the target documents.

The classification of breast cancer histopathological images study is presented by D. S. Morillo et al. in [15]. The BoF is designed based on KAZE descriptors to recognize malignant and benign tumors. The recognition based on SVM achieved 88.3 in terms of the F1-score.

In [16], R. Pal and M. Saraswat proposed a system for histopathological image classification based on enhanced BoF. The SIFT feature descriptors are extracted to produce the feature vectors. The BOF method is modified by introducing a new clustering method. The new clustering method used spiral biogeography-based optimization (SBBO) to find the optimal cluster centers. The SVM classifier is utilized to identify tissue images.

U. Muhammad et al. [17] proposed a BoF based on the KAZE descriptor for remote sensing image classification. The input image is divided into several equal blocks, and each block is treated as a separate feature extraction region. Futures selection is employed to remove useless key points. The KAZE descriptors are used to produce BoF. Then, canonical correlation analysis is utilized as a feature fusion method to refine the BOF vectors, where the fused features contain rich information and also overcome the spatial information problem. The best performance accuracy was 92 for the UCM dataset.

A. Moghimian et al. [18] proposed an image retrieval method based on multilevel BoF fusion. The image features are extracted at four levels: pixels, regions, objects, and concepts. Then, the extracted features are separately normalized. Gabor filter, SIFT, and Local Binary Pattern (LBP) descriptors are employed for feature extraction. BoF is used to describe the image at the pixel and region levels. The features are combined into a single vector using an auto-encoder network. The cosine similarity measure is used to retrieve similar images.

A method for human action recognition is presented in [19] by S. Aly and A. Sayed. First, local temporal motion energy images (MEI) are calculated. Then, local and global features are extracted using Zernike moments with different polynomial orders to represent local and global motion patterns. The whitening transformation is utilized for preprocessing global and local features. After that, BoF is used to combine these features that represent human actions. Lastly, a multi-class SVM classifier is used to recognize human actions. The best performance of the proposed method was 100% in terms of accuracy.

I. F. Nizami et al. [20] proposed a method to assess image quality. A Harris affine detector is employed to select image patches, and SIFT points are extracted over each patch. Then, the cluster centers are computed by K-means to create BoF, which are used as feature vectors. Once the features are extracted, a feature selection algorithm is utilized to select the optimal features. The selected features are provided as input to the SVM classifier to predict the image quality score. The best performance was 0.97, according to the linear correlation constant (LCC).

A. K. Shukla and S. Kanungo presented a method for face retrieval using BoF and gray wolf optimization algorithms [21]. First, the SURF feature descriptor is extracted from the face images. To produce the feature vectors, a modified BoF is utilized. BoF used the gray wolf optimization algorithm to cluster the feature vectors in place of the k-means algorithm. Finally, the face images are identified by the SVM classifier from the image dataset. The proposed method provided 96.1% overall accuracy.

R. Pal et al. [22] presented an improved BoF method for histology image classification. The predefined number of key points is detected by SURF. To reduce the computational cost of the codebook's construction, a gray relational analysis (GRA)-based key point selection method is employed to reduce the number of key points before creating visual words. The selected features are clustered by the K-means algorithm to create BoF feature vectors. The average accuracy was 78 for the performance of the proposed method.

G. Arora et al. proposed a method for diagnosing skin cancer [23]. Firstly, skin images are preprocessed to remove all types of noise. Then, the image is segmented into affected and unaffected regions. The SURF features are extracted from the segmented image, and the fusion of BoF is created. Finally, skin images are categorized into normal or abnormal skin by SVM with an accuracy of 85.7.

The image classification method is proposed by S. Vijh et al. in [24]. The modified BoF was adopted for image classification to overcome its limitations. The cat swarm optimization algorithm is employed to cluster the visual words. The weighted Gaussian mixture modeling method is used to represent the optimal visual words. The proposed classification method identified the categories of histopathological images with an accuracy of 75%.

Table 1 summarizes the related works in terms of their important topics as well as the results obtained.

Table 1: Summary of Related Works

Ref	Features	Classifier	Application	Dataset	Result	Measure
[13]	LBP and SURF	SVM	Image classification	ORL	75.0	Accuracy
				Caltech 101	79.0	
				BHSig2601	81.6	
				BHSig2602	87.0	
[14]	SIFT	SVM	Document retrieval by signatures recognition	Created by authors	99.6	Accuracy
[15]	KAZE	SVM	Breast cancer images Classification	BreaKHis	88.3	F1Scores
[16]	SIFT	SVM	Image classification	Blue ADL	69.2	Accuracy
[17]	KAZE	SVM	Image classification	NWPURESISC45	91	Accuracy
				UCM	92	
				WHU-RS	63	
[18]	Gabor filter responses and SIFT	Cosine similarity	Image retrieval	Wang	88	Precision
				Corel9C	89	
				Corel5K	68	
[19]	Zernike moments	SVM	Human action recognition	Weizmann	100	Accuracy
				KTH	84.6	
				UCF	86.4	
[20]	Harris affine and SIFT	SV M	Image quality assessment	LIVE	0.98	LCC
				TID2013	0.97	
				CSIQ	0.71	
[21]	SURF	SVM	Face retrieval	ORL	96.1	Accuracy
[22]	SURF	SVM	Histology images classification	ADL	78	Accuracy
				Blue	48	
[23]	SURF	SVM	Skin cancer recognition	PH2	85.7	Accuracy
[24]	LPB , HOG and SIFT	SVM	Histology image classification	ADL	0.75	Accuracy

The following is an explanation of the abbreviations of the terms mentioned in the previous table according to what was mentioned in the references contained therein:
ORL: Oracle Research Laboratory face database.
Caltech-101: a dataset of objects images classified into 101 classes.

BHSig2601: Bangla Signatures dataset.
BHSig2602: Hindi Signatures dataset.
BreakHis: Breast Cancer Histopathological database.
Blue: Blue histology image dataset.
ADL: Animal Diagnostic Lab dataset.
NWPU-RESISC45: a dataset of scene images classified into 45 classes.
UCM: a dataset of action images classified into 21 classes.
WHU-RS: a dataset of satellite images of google earth classified into 19 classes.
Wang: a dataset of images classified into 10 classes.
Corel9C: a dataset of images classified into 10 classes.
Corel5K: a dataset of images classified into 50 classes.
Weizmann: a dataset of action images classified into ten classes.
KTH: a dataset of action images classified into six classes.
UCF: sports dataset of realistic actions in the unconstrained environment.
PH2: a dataset of skin cancer images.
LIVE: a database of image quality assessment.
TID2013: dataset for image quality assessment.
CSIQ: Categorical Image Quality dataset for image quality assessment.

With reference to Table 1, several important observations have been clarified. Most of the proposed methods used SVM for classification, which was experimentally found to give the best performance when compared to other classifiers.

In an attempt to get a more accurate representation of the image content, the proposed methods have adopted different feature extraction techniques, whether they are used as detectors or descriptors, for example, LBP, SURF, SIFT, or HOG.

The diversity of extraction techniques and their application mechanisms was reflected in the improved performance of BoF. This is clear from the results that have been reached. However, it is difficult to determine which extraction technique is better than the others due to the difference in applications that used BOF and the difference in the dataset used. For that, this paper is designed to find the best feature extraction techniques that can help in building a BoF model according to image classification. This paper also aims to evaluate the performance of BOF when using different feature extraction techniques to facilitate comparative studies for the research community.

4. Methodology and Materials

This paper introduces an analysis method for BoF performance. Different BoF models are constructed based on different feature techniques. To achieve that, two image datasets are employed. Methodological steps are described in the following sections:

4.1 The Proposed Method of Comparison BoF Models

The purpose of the proposed method is to provide a performance analysis of different BoF models. An effective performance analysis needs to utilize different techniques of feature detection and extraction to identify the best features to improve classification performance and come up with a robust system. The proposed method is designed to find the most appropriate feature technique for constructing a BoF model. Their effects are evaluated according to image classification performance using the SVM classifier. It consists of the following major steps:

- Feature Extraction: The point detectors are employed to extract interesting points from images. Then, the features are extracted by computing descriptors for each detected point. To perform this step, several detectors and descriptors have been exploited to analyze their effect on the BoF's performance.

- BoF Model Constructing: The BoF model is formed (as explained in Section 2) to represent each image by a histogram that is built by getting the frequencies of the obtained visual word features. It is worth noting that different models of BoF are generated depending on what detectors and descriptors are used. For example, a third BoF model is created using the SURF technique as a detector and the KAZE technique as a descriptor.
- Classification Process: The SVM classifier and features obtained by the BoF model are exploited for the classification process. To conduct the experiments, image data sets are used. The data sets are partitioned into two subsets, which are training data and test data, as will be clarified in the Experimental Results section.
- Performance Evaluation: Evaluating the performance is an important step that assists in the optimization of the BoF model setting. Various measures have been used for this purpose.

4.2. Classification Datasets

For an accurate evaluation, two image datasets were adopted. Each data set has a different number of classes, which are:

4.2.1 Concrete Crack Images Dataset

Concrete Crack Images Dataset: The Concrete Crack Images dataset is collected from various METU campus buildings and generated from high-resolution images with the method proposed by [25]. The dataset contains concrete crack images for classification data. The dataset is divided into two classes: the negative class, which represents images without cracks present in the road, and the positive class, which represents images with cracks. The total number of images is 40,000, with each class containing 20,000 images. The resolution of the images is 227 x 227 pixels [26]. The dataset is publicly available at https://data.mendeley.com/datasets. Figure 2 shows examples of the types of concrete crack images in the dataset.

4.2.1 MathWorks Merch Images Dataset

MathWorks Merch Images Dataset: The MathWorks Merch dataset is an image dataset containing 75 images of MathWorks goods. The dataset is divided into five different categories: cap, cube, playing cards, screwdriver, and torch. The size of the images is 227 x 227 pixels. The dataset can be used in various applications, such as image classification and transfer learning. The dataset is available as part of the Statistics and Machine Learning Toolbox in MATLAB software. Figure 3 shows examples of data set categories.

Figure 2: Types of Concrete Crack Images in the Dataset

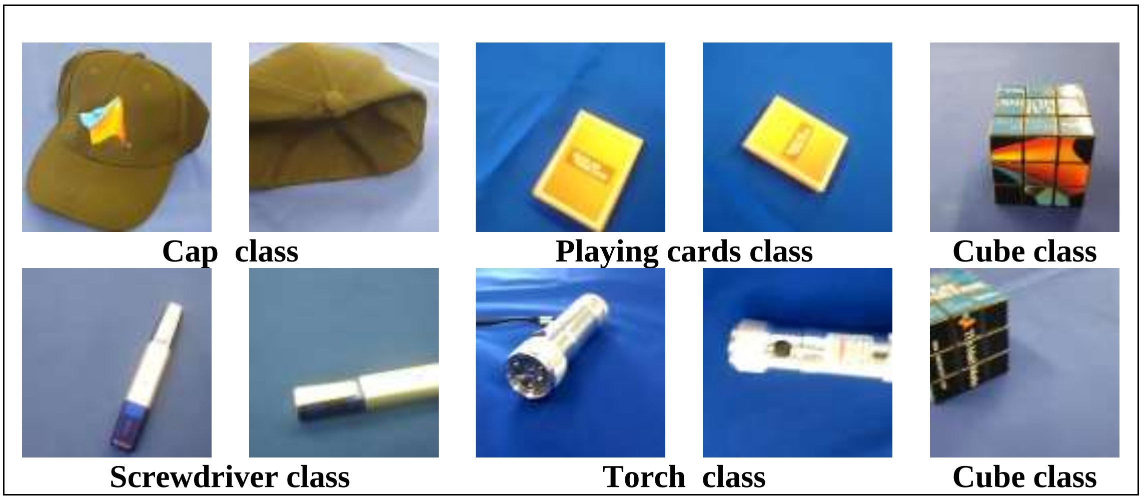

Figure 3: Class type of MathWorks Merch Images dataset

4.3. Evaluation Measures

The measure of evaluation summarizes the classification's performance. The accomplishment of the proposed method was measured by the sensitivity that gives the true positive rate, the specificity that gives the true negative rate, and the accuracy that gives the percentage of correctly classified images. They can be defined as follows [3]:

$$\text{Sensitivity}=\frac{TP}{P} \tag{1}$$

$$\text{Specificity}=\frac{TN}{N} \tag{2}$$

$$\text{Accuracy}=\frac{TP+TN}{P+N} \tag{3}$$

where P is the positive image number, N is the negative image number, TP is the number of correctly classified positive images, and TN is the number of correctly classified negative images.

5. Experimental Results Analysis

An empirical analysis of feature detectors and descriptors is presented to shed light on the most effective technique for building the BoF model. Various BoF models have been proposed with modifications to the first stage of the model (that is, the feature extraction stage). Each time a feature detector is adopted with a feature descriptor, the feature techniques used are SURF, SIFT, KAZE, and BRISK as detectors and descriptors, and HOG and LBP as descriptors. The list of proposed BoF models can be reviewed in Table 2. For example, B02 is a BoF model that uses the SURF technique as a detector and the SIFT technique as a descriptor. As for the rest of the BoF model stages, they were approved as is. The classification accuracy is analyzed using three measures and two image datasets for different proposed models of BoF. In addition, a time analysis for different models is accounted for. The first dataset (Concrete Crack) is very large, so only 2000 images are used. For classification tasks, 70% of the datasets of each class are randomly selected as the training set, while 30% of the datasets are randomly selected as the test set. This is the agreed-upon percentage for classification, according to most previous research. The method implementation and experiments are performed on a HP PC with an Intel Core i7-5500 4.40GHz CPU and 12 GB of RAM running MATLAB 2022a.

Table 2: The extracted feature methods of the proposed BoF models

BoF Models	Detector	Descriptor
B01	SURF	SURF
B02	SURF	SIFT
B03	SURF	KAZE
B04	SURF	BRISK
B05	SURF	HOG
B06	SURF	LBP
B07	SIFT	SURF
B08	SIFT	SIFT
B09	SIFT	KAZE
B10	SIFT	BRISK
B11	SIFT	HOG
B12	SIFT	LBP
B13	KAZE	SURF
B14	KAZE	SIFT
B15	KAZE	KAZE
B16	KAZE	BRISK
B17	KAZE	HOG
B18	KAZE	LBP
B19	BRISK	SURF
B20	BRISK	SIFT
B21	BRISK	KAZE
B22	BRISK	BRISK
B23	BRISK	HOG
B24	BRISK	LBP

First, the classification performance is evaluated with the concrete crack dataset according to sensitivity, specificity, and accuracy measures. The results of this experiment are presented in Table 3. Additionally, Figures 4, 5, and 6 visualize the classification performance in a graphical representation.
B15 has the highest accuracy (it is colored red), which means that the classification process is more accurate in the case of using the KAZE method as a detector and descriptor with the BoF model.

Next come B09 and B14, with comparable performances in terms of classification accuracy, followed by B08 and B13, with very close performances. The formations of the previous BoF models were based on KAZE and SIFT methods, either as detectors or descriptors. except for the B13, which depended on SURF as a descriptor. The lowest performance was provided by B12 and B24, where LBP was used as a descriptor.

Table 3: Classification performance using the concrete crack dataset

BoF Model	Sensitivity	Specificity	Accuracy
B01	0.9833	1	0.9917
B02	0.9866	1	0.9933
B03	0.9833	0.9966	0.9900
B04	0.9333	0.98	0.9567
B05	0.9533	0.99	0.9717
B06	0.3633	0.8733	0.6183
B07	0.99	1	0.9950
B08	0.9966	1	0.9983
B09	0.9966	0.9966	0.9967
B10	0.9866	1	0.9933
B11	0.9933	0.9966	0.9950
B12	0.3533	0.86	0.6067
B13	0.9966	1	0.9983
B14	0.9933	1	0.9967
B15	**1**	**1**	**1.0000**
B16	0.9633	0.9933	0.9783
B17	0.9933	0.9833	0.9883
B18	0.4133	0.8666	0.6400
B19	0.9766	0.9166	0.9467
B20	0.9	0.9966	0.9483
B21	0.91	0.9966	0.9533
B22	0.81	0.9533	0.8817
B23	0.8566	0.97	0.9133
B24	0.3533	0.86	0.6067

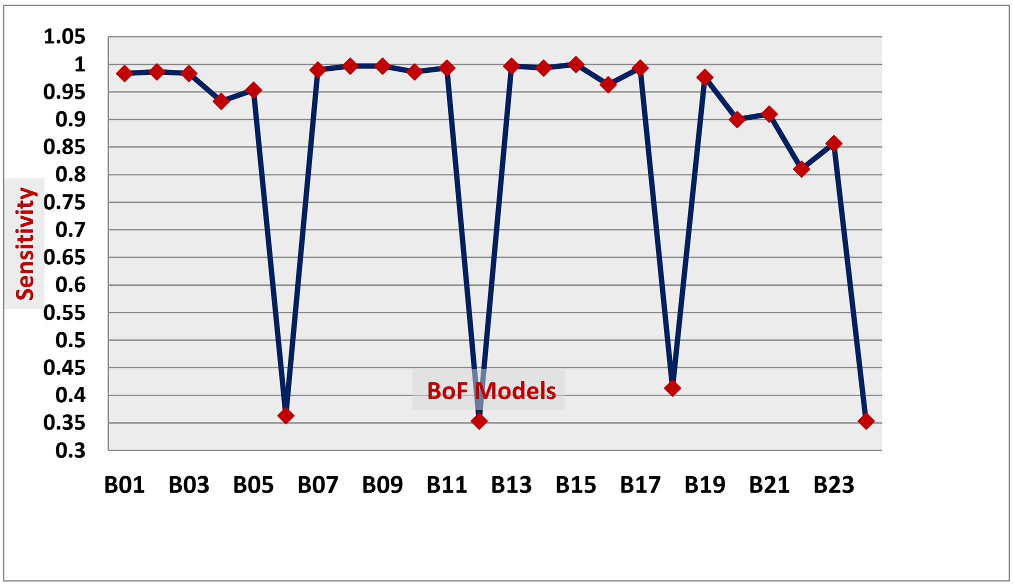

Figure 4: The results for the concrete crack dataset according to sensitivity

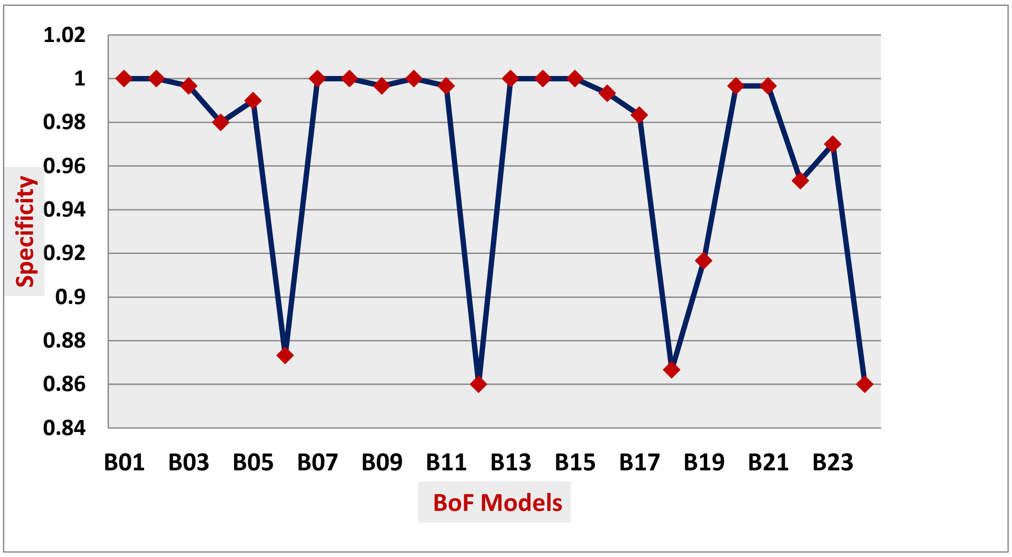

Figure 5: The results for the concrete crack dataset according to specificity

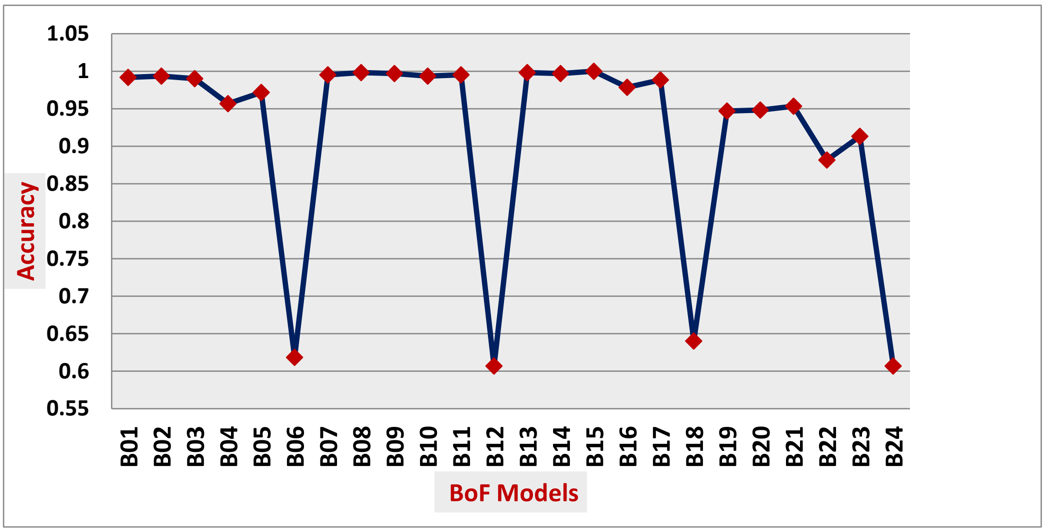

Figure 6: The results for the concrete crack dataset according to accuracy

One of the difficulties of the classification process is when the differences among instances within a single class are high and the differences among multiple classes are low. A high degree of ambiguity in various categories can greatly contribute to an error or misclassification.

In order to get more accurate results of comparison, the proposed BoF models were tested on another dataset with the above specifications. As mentioned earlier, the MathWorks Merch dataset contains five different categories that appear in different form and positions.
Table 4 summarizes the overall classification performance of the proposed BoF models using the MathWorks Merch dataset with respect to sensitivity, specificity, and accuracy measures. A clear picture was presented by visualizing the measure values as illustrated in Figures 7, 8, and 9.

Table 4: Classification performance using MathWorks' Merch dataset

BoF Model	Sensitivity	Specificity	Accuracy
B01	0.75	0.75	0.8500
B02	0.75	0.9375	0.90
B03	1	0.8125	0.85
B04	0.75	0.625	0.60
B05	0	0.5	0.60
B06	0.5	0.375	0.40
B07	0.75	0.75	0.80
B08	0.75	0.9375	0.90
B09	0.75	0.9375	0.90
B10	1	0.6875	0.70
B11	0	0.375	0.35
B12	0.5	0.375	0.40
B13	0.75	0.9375	0.90
B14	0.75	0.9375	0.95
B15	**0.75**	**1**	**0.95**
B16	0.75	0.875	0.90
B17	0.75	0.4375	0.50
B18	0.5	0.375	0.40
B19	0.75	0.75	0.70
B20	0.75	0.9375	0.85
B21	0.75	0.875	0.80
B22	0.75	0.75	0.65
B23	0.5	0.5625	0.45
B24	0.5	0.375	0.40

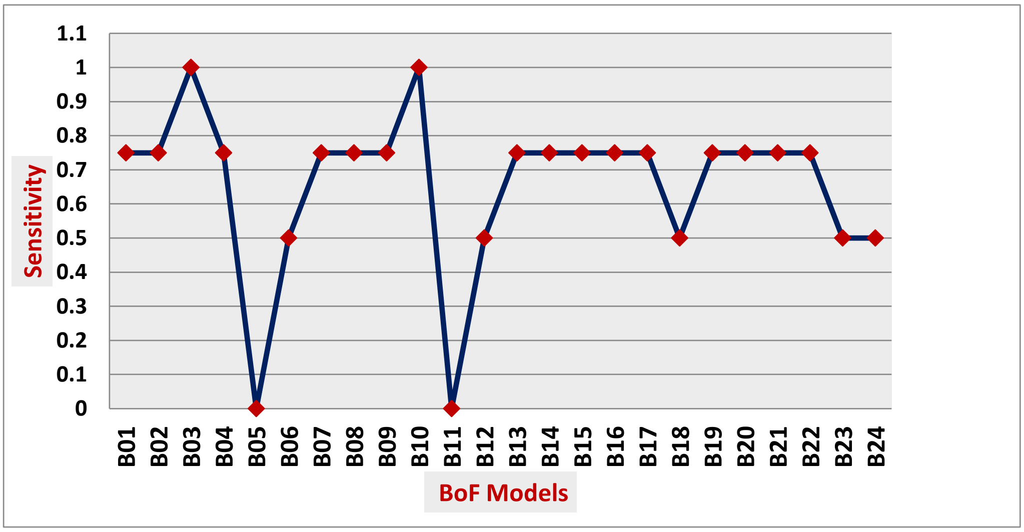

Figure 7: The results for the MathWorks Merch dataset according to sensitivity

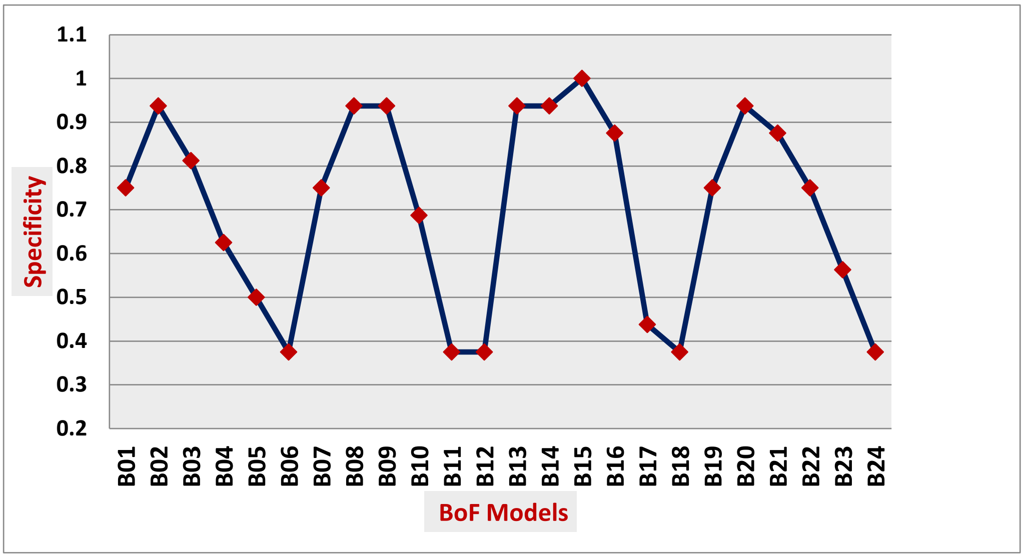

Figure 8: The results for the MathWorks Merch dataset according to specificity

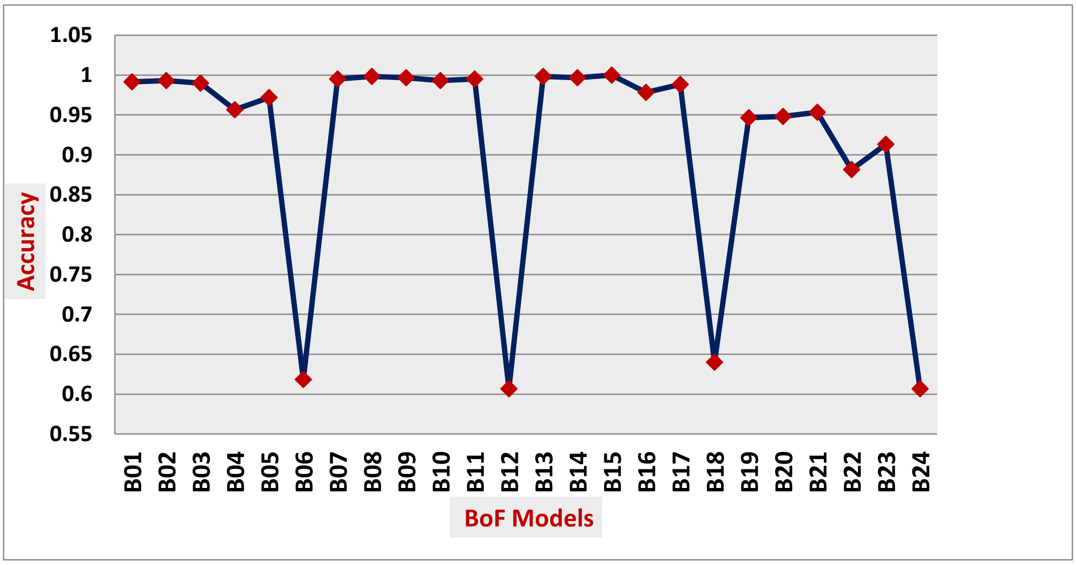

Figure 9: The results for the MathWorks Merch dataset according to accuracy

When reviewing Table 4, it seemed clear from the MathWorks Merch dataset that the use of various feature techniques has changed the classification accuracy. By comparing B15, colored in red, with the corresponding column accuracy, it is noticeable that B15 obtained a classification accuracy of 95. Again, B15 has a high classification performance level because it has the highest accuracy and specificity. It is evident that the proposed BoF model has a high classification performance when considering the KAZE method as a detector and descriptor.

One can note that the performance of the SIFT descriptor (B14) was more accurate than the SURF descriptor (B13) when KAZE was used as a detector. On the contrary, the BRISK detector achieved low accuracy. In either case, the LBP descriptor gave the lowest classification performance, as in B11.

Part of the evaluation is the time required to implement feature extraction techniques. It is an important aspect of good classification performance; therefore, this point should be focused on. For simplicity, the time comparison was accomplished in accordance with the time required for the feature extraction stage for each proposed BoF model. That is, the time required to identify the key points of the image by the detector and then describe these key points with a descriptor. Table 5 presents a comparison of execution times for different proposed BoF models. Also, the results are depicted as revealed in Figure 10.

Table 5: Execution time for proposed BoF models in seconds

BoF Model	Detector	Descriptor	Time in sec.
B01	**SURF**	**SURF**	**0.012171**
B02	SURF	SIFT	0.019925
B03	SURF	KAZE	0.019474
B04	SURF	BRISK	0.246798
B05	SURF	HOG	0.024219
B06	SURF	LBP	0.018100
B07	SIFT	SURF	0.025953
B08	SIFT	SIFT	0.038706
B09	SIFT	KAZE	0.033945
B10	SIFT	BRISK	0.257273
B11	SIFT	HOG	0.036086
B12	SIFT	LBP	0.020757
B13	KAZE	SURF	0.064885
B14	KAZE	SIFT	0.069206
B15	KAZE	KAZE	0.127430
B16	KAZE	BRISK	0.288482
B17	KAZE	HOG	0.115580
B18	KAZE	LBP	0.058262
B19	BRISK	SURF	0.243488
B20	BRISK	SIFT	0.296193
B21	BRISK	KAZE	0.257543
B22	BRISK	BRISK	0.443158
B23	BRISK	HOG	0.244461
B24	BRISK	LBP	0.230882

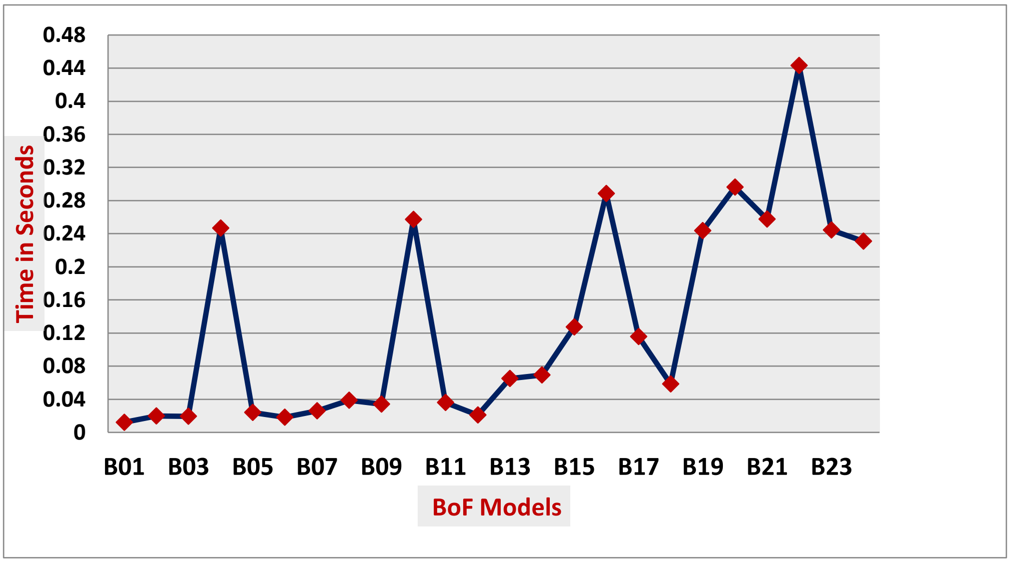

Figure 10: Execution time for proposed BoF models

The comparison was made, and it can be inferred from the time required that the application of the SURF method (B01) had the lowest execution time compared to the other proposed BoF models. Following the SURF method is the SIFT method (B08), with a slight difference of about 0.026 seconds.

On the other hand, the performance of the KAZE method was relatively low, as was evident from the time required for the B15 model. The SURF method in B01 is 10 times faster than the KAZE method in B15. Furthermore, it is noticeable that the execution time is dramatically increased when using the BRISK method, as shown in the B22 model.
Some combination possibilities, such as a SURF detector with a KAZE descriptor or a SIFT detector with a KAZE descriptor, succeeded in satisfying a high level of classification performance while maintaining a low execution time.

6. Conclusions

This paper provided a performance comparison and analysis of the BoF model. The proposed method was designed to find the best feature technique for constructing a BoF model. The proposed method first constructs BoF models based on different types of detectors and descriptors to extract the features. Then, the BoF models were evaluated based on classification performance according to sensitivity and specificity accuracy measures. The classification process was applied to two image data sets. Upon examining the results obtained, it is observed that the use of the KAZE method achieved the highest performance compared to other feature methods, whereas the use of the LBP descriptor achieved the worst performance when considering classification accuracy. On the other hand, the SURF method was superior to other methods when considering execution time. Generating a BoF model by combining the SURF or SIFT methods used as detectors and the KAZE method used as descriptors can achieve a high level of accuracy and keep the time complexity low.

Acknowledgments

The authors would like to thank Mustansiriyah University (www.uomustansiriyah.edu.iq) in Baghdad, Iraq, for its support in the present work.

References

[1] V. B. Canedo and B. Remeseiro, "Feature selection in image analysis: a survey," *Artificial Intelligence Review*, vol.53, pp. 2905–2931, 2020. https://doi.org/10.1007/s10462-019-09750-3.

[2] D. Giveki, "Scale-space multi-view bag of words for scene categorization," *Multimedia Tools and Applications*, vol. 80, pp.1223–1245, 2021. https://doi.org/10.1007/s11042-020-09759-9.

[3] N. Li and Y. Xia, "Affective image classification via semi-supervised learning from web images,"*Multimedia Tools and Applications*, vol. 77, pp. 30633–30650, 2018. https://doi.org/10.1007/s11042-018-6131-1.

[4] M. E. Abdulmunem and E. Hato, "Semantic based video retrieval system: survey," *Iraqi Journal of Science*, vol.59, no.2A, pp.739-753, 2018, doi: 10.24996/ijs.2018.59.2A.12.

[5] A. Alzu'bi, A. Amira, N. Ramzan and T. Jaber, "Improving content-based image retrieval with compact global and local multi-features," *International Journal of Multimedia Information Retrieval*, vol. 5, pp. 237–253, 2016. https://doi.org/10.1007/s13735-016-0109-4.

[6] E. Hato and M. E. Abdulmunem, "Fast algorithm for video shot boundary detection using SURF features," in *2019 2nd Scientific Conference of Computer Sciences (SCCS)*, IEEE, pp. 81-86, 2019.

[7] J. Zheng, Z. J. Wang and C. Zhu, "Food Image Recognition via Superpixel Based Low-Level and

Mid-Level Distance Coding for Smart Home Applications," *Sustainability*, vol. 9, no. 5, p. 856, May 2017, doi: 10.3390/su9050856. [Online]. Available: http://dx.doi.org/10.3390/su9050856

[8] R. S. Abdul Ameer and M. Al-Taei, "Human action recognition based on bag-of-words," *Iraqi Journal of Science*, vol. 61, no.5, pp.1202-1214, 2020, doi: 10.24996/ijs.2020.61.5.27.

[9] G. Ulutas, B. Ustubioglu, M. Ulutas and V. V. Nabiyev, "Frame duplication detection based on BoW model," *Multimedia Systems*, vol. 24, pp. 549–567, 2018. https://doi.org/10.1007/s00530-017-0581-6.

[10] J. Filali, H. B. Zghal and J. Martinet, "Comparing HMAX and BoVW Models for Large-Scale Image Classification," *Procedia Computer Science*, vol.192, pp. 1141–1151, 2021.

[11] A. A. Karim and R. A. Sameer, "Static and dynamic video summarization" *Iraqi Journal of Science*, vol.60, no.7, pp.1627-1638, 2019, doi: 10.24996/ijs.2019.60.7.23.

[12] K. S. Arun, V. K. Govindan and S. D. Madhu Kumar, "Enhanced bag of visual words representations for content based image retrieval: a comparative study," *Artificial Intelligence Review*, Springer, vol. 53, pp.1615–1653, 2020. https://doi.org/10.1007/s10462-019-09715-6.

[13] D. Srivastava, R. Bakthula and S. Agarwal, "Image classification using SURF and bag of LBP features constructed by clustering with fixed centers," *Multimedia Tools and Application*, Springer, vol. 78, pp. 14129–14153, 2019. https://doi.org/10.1007/s11042-018-6793-8.

[14] R. Mandal, P. Pratim Roy, U. Pal and M. Blumenstein, "Bag-of-visual-words for signature-based multi-script document retrieval," *Neural Computing and Applications*, vol.31, pp.6223–6247, 2019. https://doi.org/10.1007/s00521-018-3444-y.

[15] D. S. Morillo, J. González, M. G. Rojo and J. Ortega, "Classification of Breast Cancer Histopathological Images using KAZE features and Bag of Features," *In International Work-Conference of Bioinformatics and Biomedical Engineering*, Springer, pp. 204-215, 2019.

[16] R. Pal and M. Saraswat, "Histopathological image classification using enhanced bag-of-feature with spiral biogeography-based optimization," *Applied Intelligence*, vol. 49, pp. 3406–3424, 2019. https://doi.org/10.1007/s10489-019-01460-1.

[17] U. Muhammad, W. Wang, A. Hadid and S. Pervez, "Bag of words KAZE (BoWK) with two-step classification for high-resolution remote sensing images," *IET Computer Vision*, vol.13, no. 4, pp. 395-403, 2019.

[18] A. Moghimian, M. Mansoorizadeh and M. Dezfoulian, "Content based image retrieval using fusion of multilevel bag of visual words," *SN Applied Sciences*, vol. 1, Article number: 1735, pp.1-10, 2019. https://doi.org/10.1007/s42452-019-1793-5.

[19] S. Aly and A. Sayed, "Human action recognition using bag of global and local Zernike moment features," *Multimedia Tools and Applications*, vol. 78, pp. 24923–24953, 2019. https://doi.org/10.1007/s11042-019-7674-5.

[20] I. F. Nizami, M. Majid, M. ur Rehman, S. M. Anwar, A. Nasim and K. Khurshid, "No-reference image quality assessment using bag-of-features with feature selection," *Multimedia Tools and Applications*, vol. 79, pp. 7811–7836, 2020. https://doi.org/10.1007/s11042-019-08465-5.

[21] A. K. Shukla and S. Kanungo, "Automated face retrieval using bag‑of‑features and sigmoidal grey wolf optimization," *Evolutionary Intelligence*, vol.14, pp. 1201–1212, 2021. https://doi.org/10.1007/s12065-019-00213-w.

[22] R. Pal, M. Saraswat and H. Mittal, "Improved bag-of-features using grey relational analysis for classification of histology images," *Complex & Intelligent Systems*, vol. 7, pp. 1429–1443, 2021. https://doi.org/10.1007/s40747-021-00275-3.

[23] G. Arora, A. K. Dubey, Z. Abdin Jaffery and A. Rocha, "Bag of feature and support vector machine based early diagnosis of skin cancer," *Neural Computing and Applications*, Springer, vol. 34, pp.8385–8392, 2022. https://doi.org/10.1007/s00521-020-05212-y.

[24] S. Vijh, S. Kumar and M. Saraswat, "New bag-of-feature for histopathology image classification using reinforced cat swarm algorithm and weighted Gaussian mixture modeling," C*omplex & Intelligent Systems*, vol. 8 , pp. 5027–5046, 2022. https://doi.org/10.1007/s40747-022-00726-5.

[25] L. Zhang, F. Yang, Y. D. Zhang and Y. J. Zhu, "Road Crack Detection Using Deep

Convolutional Neural Network," *In International Conference of Image Processing (ICIP)*, IEEE, pp.3708-3712, 2016.

[26] R. Fan, M. Junaid Bocus, Y. Zhu, J. Jiao, L. Wang, F. Ma, S. Cheng and M. Liu, "Road Crack Detection Using Deep Convolutional Neural Network and Adaptive Thresholding," I*n International Conference of Intelligent Vehicles Symposium (IV)*, IEEE, pp. 474-479, 2019.