

# A Comparative Study of Data Mining approaches for Bag of Visual Words Based Image Classification

Sahar Abd El Hameed Mokhtar and Heba Ahmed Elnemr

Department of Computers and Systems, Electronics Research Institute, Giza, Egypt

## Article history

Received: 08-10-2017

Revised: 11-12-2017

Accepted: 11-01-2018

Corresponding Author:

Heba Ahmed Elnemr

Department of Computers and  
Systems, Electronics Research  
Institute, Giza, Egypt

Email: heba@eri.sci.eg

heba\_elnemr@yahoo.com

**Abstract:** Image classification is one of the most significant and challenging tasks in computer vision. The goal of this task is to build a system that is capable to reveal an image label within a collection of different image categories. This paper presents and discusses the application of various data mining techniques for image classification based on Bag of Visual Words (BoVW) feature extraction algorithm. The BoVW model is constructed using grey level features: The Speeded Up Robust Features (SURF) and Maximally Stable Extremal Regions (MSER) descriptors along with color features: Color correlograms and Improved Color Coherence Vector (ICCV). Five data mining techniques; Neural Networks (NN), Decision Trees (DT), Bayesian Network (BN), Discriminant Analysis (DA) and K Nearest Neighbor (KNN), are explored and evaluated on two large different datasets: Corel-1000 and COIL-100. The experimental results illustrate that BN and DA outperform the other data mining methods considered in this comparative study. For Corel-1000 dataset, BN and DA achieved an average accuracy and specificity of about 99.9% and an average sensitivity of about 99.5 and 99.4%, respectively. While for the COIL-100 dataset, BN and DA accomplished an average accuracy and sensitivity of about 100% and an average specificity of about 98.5 and 98.9, respectively.

**Keywords:** Image Classification, Data Mining, Machine Learning, Neural Networks, Decision Trees, Bayesian Network, Discriminant Analysis, K Nearest Neighbor

## Introduction

Image classification is one of the significant and challenging research areas in the fields of computer vision and pattern recognition. It is a key task in many application domains, including image and video retrieval, document image classification, scene understanding, video surveillance, remote sensing, robot navigation, vehicle navigation, biomedical imaging, biometrics, etc.

Image classification is the process of assigning images to one of a number of predefined semantic categories using their features. Image classification procedures can be assorted into two main groups, namely supervised classification and unsupervised classification. In supervised classification, the set of classes is identified beforehand. Though, in unsupervised classification, the set of likely classes are not identified. The supervised image classification is carried out by first obtaining statistical characterizations of predefined

informational classes through a supervised training step. Then, the image is classified by examining its statistical characterization and deciding about which of the classes it seems like most. Unsupervised classification, on the other hand, has no training step to find the target class. Unsupervised image classifiers examine great numbers of unidentified pixels of different images and group them into a number of clusters based on their characteristics (Kamavisdar *et al.*, 2013).

Image classification is a complicated process that may be influenced by many factors; the appearance of images of the same category can vary substantially due to imaging and lighting conditions alterations, poses changes, within-category structure variations and the presence of noisy or blurry contents (Lu and Weng, 2007; Jain and Tomar, 2013; Kamavisdar *et al.*, 2013; Naswale and Ajmire, 2016). Generally, the image classifier should be flexible enough to categorize a wide range of visually distinct classes, each with large within-

class differences, while still preserving good discriminative ability among different classes. To attain this objective, several machine learning and data mining techniques are implemented (Spangler *et al.*, 1999).

In this study, we demonstrate an inclusive experimental comparative study of the use of 5 popular data mining techniques for multi-class image classification. In general, image classification is a two-step procedure; feature extraction followed by classifier implementation. In the feature extraction step, we applied the feature extraction algorithm presented in (Elnemr, 2016). This algorithm is based on extracting image keypoints and then mapping these keypoints into visual words. Recently, there is a tendency of applying image keypoints (interesting points) or keyregions (interesting regions) for retrieving and classifying images (Elnemr, 2016; Yang *et al.*, 2007; Liu *et al.*, 2016; Wang *et al.*, 2016; Benkrama *et al.*, 2013). Keypoints and keyregions contain valuable local information about an image that can be automatically detected by exploiting several detectors and characterized by many descriptors (Lowe, 2004; Krig, 2014). The obtained descriptors are then gathered into a large number of clusters such that similar descriptors are assigned to the same cluster. Each cluster is treated as a visual word that signifies a specified local pattern shared by the descriptors in that cluster. As a result, a visual-word vocabulary portraying all types of such local image patterns is created. Thus, an image can be characterized as a vector holding the count of each visual word in that image that is utilized as a feature vector in the image classification task. The Speeded Up Robust Features (SURF) and Maximally Stable Extremal Regions (MSER) descriptors are combined with color correlograms and Improved Color Coherence Vector (ICCV) feature vectors and used to build the BoVW Model. The generated visual word vectors are fed to the Neural Networks (NN), Decision Trees (DT), Bayesian Network (BN), Discriminant Analysis (DA) and K Nearest Neighbour (KNN) to perform the classification task. The performances of these classifiers are examined and evaluated on a large collection of benchmark dataset.

The paper is organized as follows: A short overview of machine intelligence and data mining and a brief overview of data mining techniques are presented in section 2, while section 3 reviews some related work on image classification. The procedure for the comparative study is exhibited in section 4. The experimental results are discussed in section 5. The conclusion is provided in at the end.

## Machine Learning and Data Mining

Machine Learning (ML) is to search in a very large space of possible concepts to choose one that best fits the

observed data and any prior knowledge held by the learner (Mitchell, 1997). ML techniques are algorithms for discovering patterns in data as structural descriptions from examples. It can be used to predict outcome in a new situation. Methods of ML originate from artificial intelligence, statistics and research on databases (Witten *et al.*, 2011). Machine learning could be divided into supervised learning that includes classification techniques and unsupervised learning that includes clustering techniques.

Data mining is defined as the process of discovering patterns in data, using machine learning techniques for helping to explain that data and make predictions from it. The data will take the form of a set of examples, or situations. The output takes the form of predictions about new examples or prediction of a particular class or category. As well as performance, it is helpful to supply an explicit representation of the knowledge that is acquired.

Many learning techniques look for structural descriptions of what is learned, these descriptions serve to explain what has been learned. Four basically different categories of learning appear in data mining applications. In classification learning, the learning strategy is presented with a set of classified examples of a concept from which it is expected to learn a way of classifying unseen examples. In association learning, any association among features is considered, not just ones that predict a particular class value. In clustering, groups of examples that belong together are sought. In numeric prediction, the outcome to be predicted is a numeric quantity, not a discrete class. Regardless of the type of learning involved, we call the thing to be learned the concept and the output produced by a learning scheme the concept description (Witten *et al.*, 2011).

Data mining for image classification is an essential research area in computer science. It is a very effective and challenging task in several application domains, including medicine (Dash and Panda, 2016; Diz *et al.*, 2016; Singh *et al.*, 2016), remote sensing (Lu and Weng, 2007), facial micro expressions (Huang *et al.*, 2012; 2013; 2016), face recognition (Luo and Zhang, 2014) and etc.

In this study, different data mining techniques are applied to the image classification problem and comparative results are presented and discussed.

### *Data Mining for Image Classification*

Nowadays, data mining is becoming an exemplary tool to efficiently analyze and classify large image datasets.

In this section, we are going to represent the most frequently used data mining techniques for image classification, namely NN, DT, BN, DA and KNN classifiers.

### *Neural Networks Classifier (NN Classifier)*

NN classifier has emerged as an effective technique in data mining for classifying images. The terminology

neural network is manifested of the human neural network. NN techniques are data-driven and self-adaptive schemes that can adjust themselves to the data, without any specific description of the underlying model function. Furthermore, they are considered universal functional approximations since they can approximate any function with an arbitrary accuracy. Basically, NN contains three kinds of layers, input layer, hidden layers and output layer. Data goes into the network through the input layer, then fed through the hidden layers to protrude from the output layer. Each layer consists of a group of neurons. At each layer, the neurons are interconnected with weighted connections. These weights are automatically adjusted during the training procedure (Han and Kamber, 2006; Abe *et al.*, 2012; Zhang, 2000).

The network learns by examining individual example, generating a prediction for each example and making adjustments to the weights whenever it makes an incorrect prediction. This process is repeated many times and the network continues to improve its predictions until one or more of the stopping criteria have been met. In the beginning, all weights are random and the answers that come out of the net are probably nonsensical. The network learns through training. Examples for which the output is known are repeatedly presented to the network and the answers it gives are compared to the known outcomes. Information gathered from this comparison is drawn back through the network, gradually adjusting the weights. As training progresses, the network becomes increasingly accurate in replicating the known outcomes. Once trained, the network can be applied to future cases where the outcome is unknown. An NN can approximate a wide range of predictive models with minimal demands on model structure and assumption. (IBM SPSS, 2015). In spite of its flexibility, the neural network is not easily interpretable.

In this study, we focused on MultiLayer Perceptron (MLP) that is composed of three layers: Input layer, output layer and one hidden layer. MLP uses backpropagation algorithm to learn its connection weights. This method is characterized by its robustness and simplicity (Witten *et al.*, 2011; IBM SPSS, 2015; Han and Kamber, 2006).

#### *Decision Tree Classifier (DT Classifier)*

DT classifier is a nonparametric data mining technique, which is represented by a tree architecture with numerous branches and leaves. Each node signifies a test for a certain attribute, each branch denotes an outcome of the test and tree leaves enclose the predicted classes. There are three types of nodes, including the root, the internal nodes and the leaf nodes. A decision is determined using a hierarchical rule-based method that selects the path to be followed, starting from the root and passing through successive internal nodes

until a leaf node is attained, which represents the class of the image being categorized (Tan *et al.*, 2004; Han and Kamber, 2006).

An advantage of DT classifier is that it can manipulate high dimensional training data and it does not need a massive design and training. Also, it has a simple structure, which permits interpretation and visualization.

In order to overcome tree overfitting, either one of two basic strategies may be applied. The first strategy is to halt the expansion of the tree when certain criterion has been encountered, while the second strategy is achieved after the DT has been built by iteratively integrating leaf nodes to reduce large trees (tree pruning) (Han and Kamber, 2006). Different models of the Decision tree are used in classification problems, they differ in their splitting algorithms that maximize the purity of the resulting classes of data samples. Examples of the different models of the decision tree are ID3, Quinlan, C4.5, CHAID and C5.0. (Fakhr and Elsayad, 2012).

In this study, the C5.0 decision tree model is used, which is splitting the training sample based on the field that provides the maximum information gain (Han and Kamber, 2006; IBM SPSS, 2015). Each subsample determined by the initial split is then divided again, commonly based on another field and this procedure is repeated until the subsamples cannot be split again. Finally, the lowest-level splits are re-examined and those that do not contribute significantly to the value of the model are removed or pruned.

#### *Bayesian Network Classifiers (BN Classifiers)*

The Bayesian classifier is a statistical classifier that is based on Bayes' theorem. A Bayesian classifier constructs a probabilistic model, which uses the posterior probabilities to predict the class label of a tested sample. The classifier applies Bayes theorem to estimate the posterior probability from the prior probability that is computed from features gathered from the training samples. Bayesian classifiers have also demonstrated high accuracy and speed when applied to huge databases (Murty and Devi, 2011).

BN comprises a network of nodes, one for each attribute, linked by directed edges with no cycles, which is called a directed acyclic graph. A probability model is built using the BN that combines known evidence with "common-sense" real-world information in order to represent the likelihood of occurrences using apparently unlinked attributes. The main advantage of a BN is that it is a graphical model that displays variables (often referred to as nodes) in a dataset and the probabilistic, or conditional, independencies between them. Causal relationships between nodes may be presented by a BN; however, the links in the network (known as arcs) do not necessarily represent the direct cause. Networks are very robust where information is

missing and make the best possible prediction using whatever information is present.

The BN helps the user to learn about causal relationships among different features and clauses since it enables to realize the problem area and to foretell the effects of any interference. Furthermore, the network provides an efficient approach for avoiding the overfitting of data. Besides, a clear visualization of the relationships involved is easily observed.

There are two methods for constituting of BN models based on the Naive Bayes model (IBM SPSS, 2015):

- Tree Augmented Naïve Bayes. It effectively creates a simple BN model. The model is an improvement over the naïve Bayes model as it allows for each predictor to depend on another predictor in addition to the target variable. Its main advantages are its classification accuracy and favorable performance compared with general Bayesian network models. Its disadvantage is also due to its simplicity; it imposes many restrictions on the dependency structure uncovered among its nodes
- Markov Blanket estimation. The Markov blanket for the target variable node in BN is the set of nodes containing target's parents, its children and its children's parents. Markov blanket identifies all the variables in the network that are needed to predict the target variable. This can produce more complex networks but also takes longer to produce

In this study, the Tree augmented method is applied.

#### *Discriminant Analysis Classifier (DA classifier)*

DA is a multivariate statistical scheme that derives a linear equation to combine independently observed attributes or predictor variables, which discriminate effectively among the classes. The linear combination of these quantitative variables is known as the discriminant function. The discriminant function generates both raw and standardized coefficients that can be used as weights to discover the best attributes to contribute in discriminating among dependent groups (Fernandez, 2002; Ramayah *et al.*, 2010).

DA builds a predictive model for group membership. The model is made up of a discriminant function (or a set of discriminant functions for more than two groups) that is based on linear combinations of the predictor variables, which provide the best discrimination between the groups. The functions are produced from a sample of instances for which group membership is recognized. Afterward, these functions can be applied to new instances that have measurements for the predictor variables with an unknown group membership (IBM SPSS, 2015).

#### *K Nearest Neighbour Classifier (KNN Classifiers)*

Nearest Neighbours Analysis is a method for classifying cases based on their similarity to other cases. This method is called instance-based learning. In machine learning, it was developed as a way to recognize patterns of data without requiring an exact match to any stored patterns, or cases. Cases that are near each other categorized as similar cases while cases that are distant from each other are categorized as dissimilar cases. Thus, the distance between two cases is a measure of their dissimilarity. Neighbours are instances of a class that are near each other. When a new example is seen, its classification is determined by calculating its distance from each instance in the model. The most similar cases - the nearest neighbours - are classified, then the new case is located in the category that holds the greatest number of nearest neighbours (Witten *et al.*, 2011; IBM SPSS, 2015). The difference between this method and the others the time at which the "learning" takes place. The learning is lazy, deferring the real work as long as possible, whereas other methods are eager, producing a generalization as soon as the data has been seen. Sometimes more than one nearest neighbour is used and the majority class of the closest K neighbours (or the distance-weighted average if the class is numeric) is assigned to the new instance. This is termed the k-nearest-neighbor method. In this study, the KNN classifier is used with K between 3 and 5 and the calculation method of distance is Euclidian metric.

#### **Related Work**

In this section, we display some reported studies in image classification.

Luo and Zhang (2014) offered a hybrid approach for image classification that combines the Extreme Learning Machine (ELM) and the Sparse Representation based Classification (SRC) methods. The suggested approach is applied to handwritten digit image classification and face recognition.

The work of Diz *et al.* (2016) presented a data mining based approach for breast cancer classification and diagnosis. The suggested method proceeds by first obtaining two feature-based matrices; the Gray-Level Co-occurrence Matrix (GLCM) and the Gray-Level Run Length Matrix (GLRLM). Then, various classification methods are implemented, including, k-nearest Neighbour, Support Vector Machines, Decision Trees and Naive Bayes. Waikato Environment for Knowledge Analysis (WEKA) is used to apply the data mining tasks.

Austin *et al.* (2013) compared the performance of various classification techniques to classify patients with heart failure subtypes. The comparison is carried out among the conventional classification trees and the classification schemes developed in data-mining and

machine-learning literature, including bootstrap aggregation (bagging), boosting, random forests and support vector machines.

A classification model to classify benign and malignant tumors in breast ultrasound images is proposed by (Singh *et al.*, 2016). The model is based on combining fuzzy c-means clustering and back-propagation artificial neural network. In this investigation, a total of 457 features including of 447 texture and 10 shape features are extracted from the breast ultrasound images. Then, Multiple traditional state-of-art feature selection techniques are used and evaluated to choose the most relevant features. Finally, the fuzzy c-means clustering and back-propagation artificial neural network are combined to build the classification model. The performance of the suggested technique is compared with that of the back-propagation artificial neural network and the support vector machine along with some of the recently stated searches.

The usage of discriminant analysis for multi-class classification is explored by (Li *et al.*, 2006). The performance of discriminant analysis is assessed on a large collection of benchmark datasets, besides its usage in text categorization is investigated.

The work of (Acosta-Mendoza *et al.*, 2012) applied a subgraph mining algorithm for image classification. Firstly, a graph-based image representation is obtained. Afterward, the Frequent Approximate Subgraph (FAS) miners are utilized to get all the FAS patterns from the graph collection. These obtained FASs are deemed analogous to the vocabulary procured in the bag-of-features method. Therefore, each image is represented by feature vectors that are built from those FASs as features. Finally, the feature vectors are fed to an SVM classifier.

The authors of (Liu *et al.*, 2016) introduced the use of a Bag-of-Structural Words (BoSW) model that encodes the spatial attributes between a pair relevant points, for image classification. At first, the interesting points are obtained and quantized to build the Bag-of-Visual Words (BoVW). Then, a structure feature is computed for each pair related points to create the BoSW that bestows structural information through the interest points. Finally, the codebook histograms of BoVW and BoSW are combined and used to train an SVM classifier.

Dash and Panda (2016) implemented three widespread data mining techniques, including Naive Bayes, Decision trees and Random Forest algorithm for classification of medical, satellite and scenery images freely available on the Internet. The experiments are carried out for normal, noisy and filtered images with the three data mining algorithms to check their efficiency regarding the different image models.

The authors in (Wang *et al.*, 2017) present a comparative study of image classification techniques for automatic diagnosing ophthalmic diseases. In this study, typical methods for feature extraction were combined

with various classification techniques in different schemas to identify ophthalmic diseases. The performances of these schemas were compared regarding multiple aspects. The gray tone spatial dependence matrices, gray gradient co-occurrence matrices, Wavelet transformation, Local Binary Pattern (LBP) and sparse representation were used for feature extraction. While, Extreme Learning Machine (ELM), SVM, Genetic Algorithm (GA), kNN and Differential Evolution (DE) are used for classification.

The work of (Hosseini and Kandovan, 2017) proposed a hierarchical algorithm for hyperspectral image classification based on SVM. The classification task is accomplished, by the suggested hierarchical algorithm, through two levels. In the world-class level, clusters that comprise similar classes are delineated by the computing Euclidean distance between the class centers. The SVM algorithm is achieved on the clusters with selected features. Next, the classes in every cluster are separated based on SVM technique and less features. Correlation criteria between classes and features is used to select the features in every level.

## Procedure for the Comparative Study

This section states the procedure for fulfilling the comparative study of various data mining techniques for image classification. Figure 1 portrays the procedure of the comparative study.

### Strategy of the Comparative Study

To compare different classification techniques, we retain the feature extraction methodology fixed and only change the classification approaches within the whole procedure. In this study, we implemented BoVW model for obtaining the discriminative features for images owing to its simplicity, robustness and efficiency. In BoVW model, local features are extracted and then visual vocabulary is built for each training class.

### Bag-of-Visual-Words (BoVW) Model

The bag-of-words model has been comprehensively applied the technique for text categorization, which depicts the document as a set of distinctive keywords. The BoVW scheme is analogous to the bag-of-words technique. Due to the simplicity and good performance of the BoVW methodology, it has been recently applied to image classification by treating image features as words. The BoVW algorithm, generally, proceeds in three steps. The first step involves detecting and extracting local features. The second step comprises constructing the codebook (visual word vocabulary) by clustering all the obtained features. The codebook can be considered as a dictionary that records corresponding mappings among features and their description in the image. In the third step, the frequency of each visual word in the image is computed.

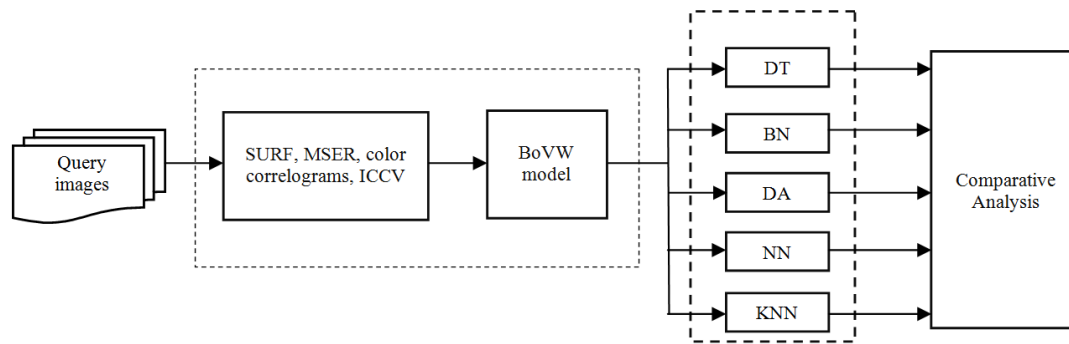


Fig. 1: Procedure of the comparative study

In view of that, BoVW scheme creates a histogram of visual features occurrences that represents an image (Liu *et al.*, 2016; Elnemr, 2016). The obtained histograms are utilized to train an image category classifier.

In order to build the dictionary of visual words, feature extraction task is performed. In this work, we implemented the feature extraction approach proposed by (Elnemr, 2016). The algorithm is carried out by detecting SURF interesting points and MSER interesting regions. Afterward, distinctive feature descriptors are computed for each point and region. As SURF and MSER operate only on grey scale images, color correlograms and ICCV are implemented to extract color features.

SURF is a pioneering scale and rotation invariant key point detector and descriptor that was initially launched by (Bay *et al.*, 2006). SURF outperforms previously proposed detectors and descriptors besides its computation is extensively very fast. For each image, SURF algorithm engenders a set of interesting points, which are represented by a set of 64-dimensional descriptors for each.

Moreover, MSER is a feature detector method that extracts elliptical covariant regions from images based on watershed algorithm (Matas *et al.*, 2002). These regions are considered as the interesting regions and a set of descriptors are extracted from each region using SURF technique. Therefore, for each image, there is a set of interesting regions that each encloses a set of key points. Each key point is represented by 64 -dimensional descriptors. MSER technique has several advantages include; features are variable size and are computed globally across an entire region, not limited to patch size or search window size and MSER regions are invariant to affine transform.

On the other hand, color correlograms (Huang *et al.*, 1997) and ICCV (Chen *et al.*, 2007) are applied to extract the color features. Color correlograms feature signifies the color correlation in an image as a function of their corresponding spatial distances. It has the advantage of capturing the pixels color distribution as color histogram along with their spatial information in the image. The size of color correlograms feature swivels on the number of quantized colors employed for feature extraction. In this exertion, the RGB color model is applied and 64 quantized

colors with two distances are implemented. Therefore, the correlograms feature vector is of size  $2 \times 64$ .

Moreover, to extract ICCV feature vector, the color histogram is divided into two constituents: A coherent constituent that holds pixels that are spatially connected and a non-coherent constituent that includes pixels that are separated. Besides, ICCV encloses more spatial information than that of conventional color coherence vector, which enhances its performance without much extra computation (Chen *et al.*, 2007). In this investigation, the ICCV feature vector is depicted of 64 coherence pairs, each pair characterizes the number of coherent and non-coherent pixels of a particular color in the RGB space. Accordingly, the ICCV feature vector is of size  $2 \times 64$ .

Each image is represented by a set of keypoint descriptors. However, these sets vary in cardinality which produces difficulties in learning techniques (classifiers) that require feature vectors to be of the same dimension. Thus, there is an urge to quantize the extracted local descriptors in their feature space into visual words (clusters) to form the visual dictionary. K-means clustering algorithm is employed to build the visual word vocabulary that describes the different local patterns in the images. The number of clusters determines the vocabulary size.

As a result, an image can be portrayed as a vector of words the same as a document. For each image, a BoVW (histogram of words) is built holding the occurrences of each individual word in the dictionary of that image, which will be used as a feature vector in the classification stage.

### Classification Procedure

After constructing the BoVW model, it is implanted into the classification phase for training and testing. In this study, various data mining techniques were put through for image classification, so as to get the best classifier that is best fitted to our problem. It is worth noted that no data mining method can be considered better than others since each method has its cons and pros. Therefore, we test several data mining techniques and evaluate them. For this review NN, DT, BN, DA and KNN classifiers are considered.

## Comparative Analysis

This section explains in details the examined datasets, the used evaluation strategy and the comparative results for the implemented data mining algorithms. This comparative study portrays the objectives and drawbacks of the applied data mining algorithms.

### Dataset

The comparative study was evaluated using two different benchmark datasets; Columbia Object Image

Library (COIL-100) dataset (Nene *et al.*, 1996) and a subset of the Corel image database (Wang *et al.*, 2001).

COIL-100 is a popular database of color images that involves 100 different objects. Each object was acquired with different viewing angles by rotating the object around the vertical axis to obtain 72 views.

The utilized Corel-1000 database consists of 10 irrelevant arbitrary real word classes, each class holds 100 color images from the Corel stock photo database. The database contains Dinosaur, Cyber, Horse, Bonsai, Texture, Fitness, Dishes, Antiques, Elephant and Easter egg groups. Figure 2 and 3 illustrate samples of these datasets.



Fig. 2: Sample images from COIL-100 database



Fig. 3: Sample images from Corel-1000 database (a) Dinosaurs, (b) Cyber, (c) Horses, (d) Bonsai, (e) Texture, (f) Fitness, (g) Dishes, (h) Easter egg, (i) Antiques, (j) Elephants

### Evaluation Methods

The evaluation of classification methods is one of the most lucrative topics of experimental analysis that permits objectively selection of the appropriate method for a given data. Usually, confusion matrix, as well as three statistical measurements: Accuracy, sensitivity and specificity are used to assess the system's performance.

A confusion matrix is a cross-tabulation of your predicted values against the true observed values and (test) accuracy is an empirical rate of correct predictions.

The confusion matrix is a table that is commonly applied to visually depict the classifier performance on a set true recognized test data. Each column (row) of the matrix denotes the predicted classes of the input instances while each row (column) denotes the actual classes of the input. Four basic terms can be obtained from the confusion matrix, namely:

- True Positives (TP): Positive instances that are correctly predicted
- True Negatives (TN): Negative instances that are correctly predicted
- False Positives (FP): Positive instances that are incorrectly predicted
- False Negatives (FN): Negative instances that are incorrectly predicted

On the other hand, the accuracy is an empirical rate of correct prediction, the sensitivity is the ability to correctly classify images as belonging to a particular class and the specificity is the ability to predict that images of other classes are not part of a stated class. These performance measures are computed as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

### Comparative Results

This study aims to evaluate and compare five data mining methods naming NN, DT, BN, DA and KNN classifiers, for image classification. These methods are trained and tested using the BOVW model proposed in (Elnemr, 2016).

The accuracy, sensitivity and specificity of the data mining algorithms in image classification, for each class in the Corel-1000 dataset, are illustrated in Fig. 4 to 6, respectively. From Fig. 4 we can notice that the highest accuracy achieved is 100% using a Bayesian classifier for dinosaurs, horses, bonsai and fitness classes, as well as using the DA classifier for the bonsai and antique classes. Furthermore, Fig. 5 illustrates that the maximum sensitivity attained is 100% using a Bayesian classifier for categories dinosaurs, cyber, horses, bonsai, fitness, antiques and elephants and using a DA classifier for categories bonsai, textures, fitness, dishes, Easter egg and antiques. Likewise, the specificity reached 100% when applying a Bayesian classifier for classes' dinosaurs, horses, bonsai, fitness, dishes and Easter egg, besides a DA classifier for classes' dinosaurs, cyber, horses, bonsai, antiques and elephants (Fig. 6).

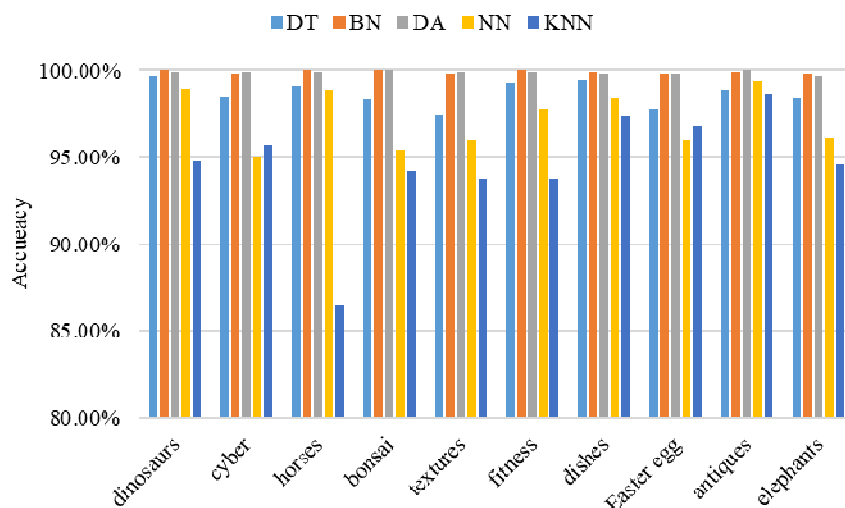


Fig. 4: Accuracy value for different classes of Corel-1000 dataset using different data mining techniques



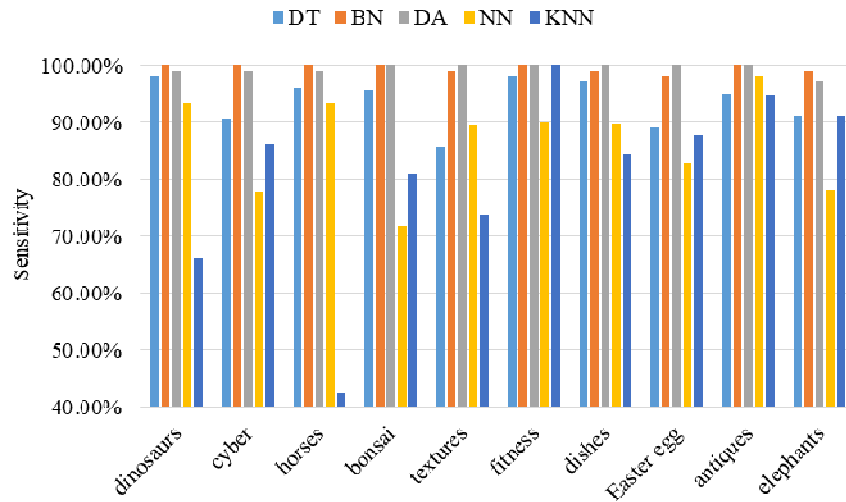


Fig. 5: Sensitivity value for different classes of Corel-1000 dataset using different data mining techniques

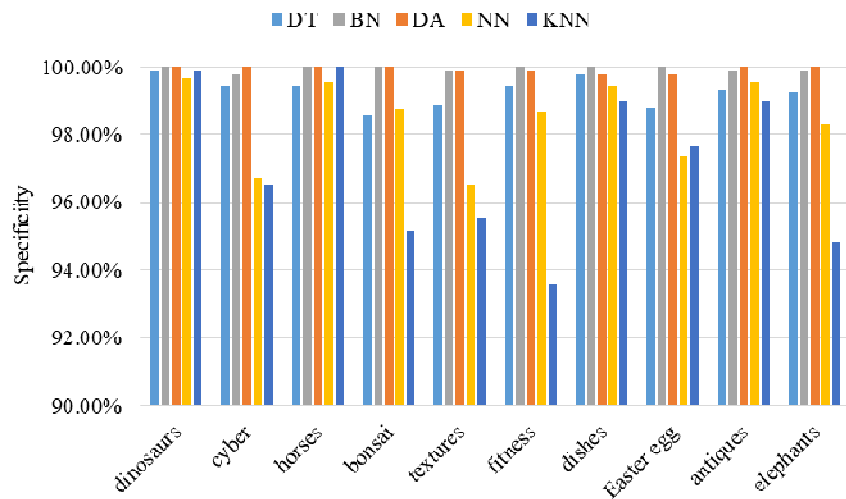


Fig. 6: Specificity value for different classes of Corel-1000 dataset using different data mining techniques

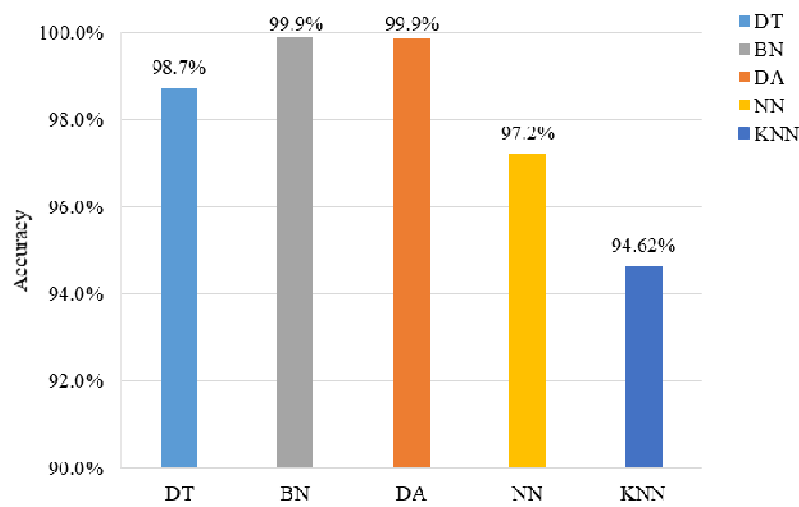


Fig. 7: Average accuracy value for Corel-1000 dataset using different data mining techniques

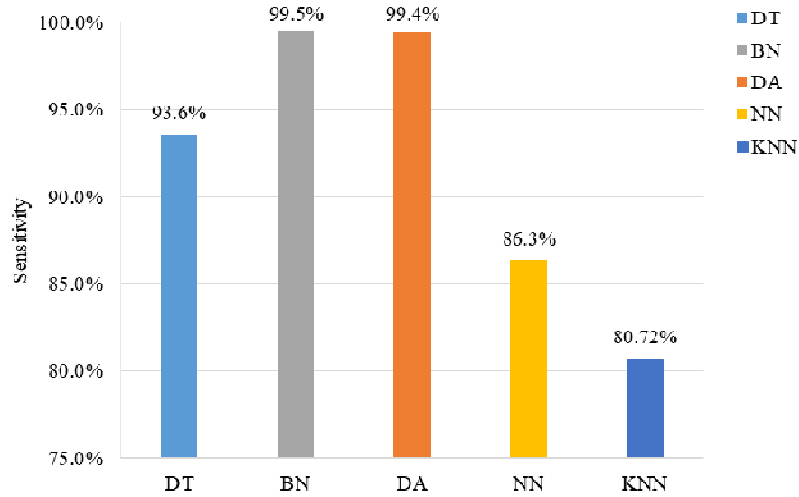


Fig. 8: Average sensitivity value for Corel-1000 dataset using different data mining techniques

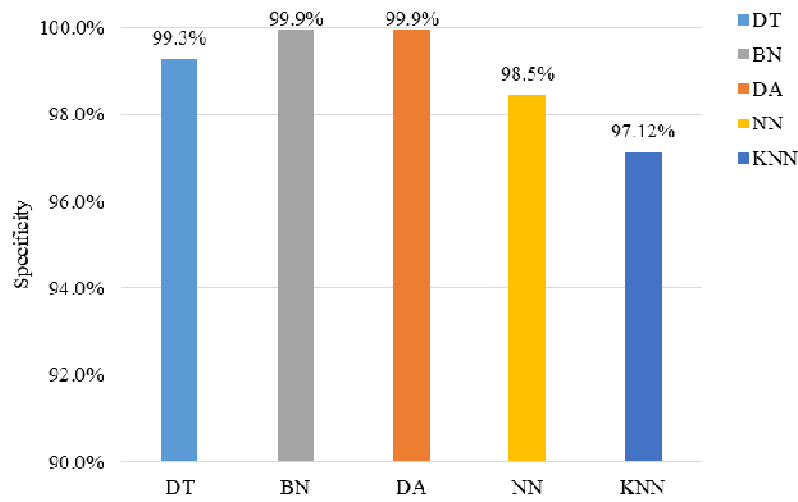


Fig. 9: Average specificity value for Corel-1000 dataset using different data mining techniques

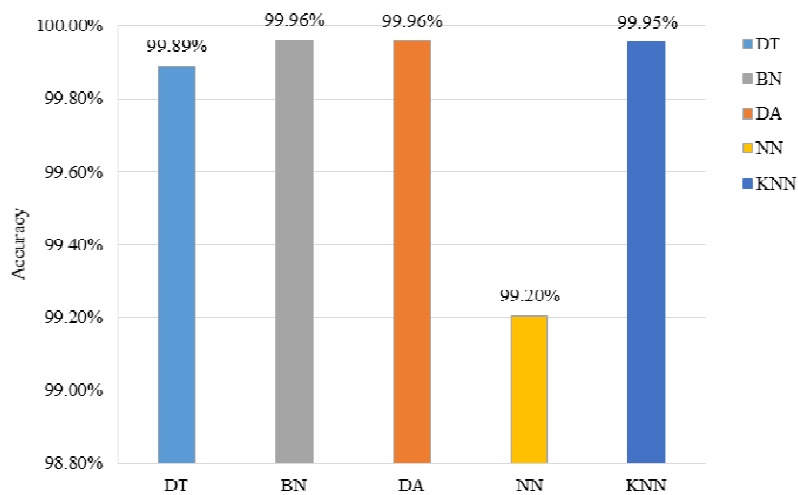


Fig. 10: Average accuracy value for COIL-100 dataset using different data mining techniques

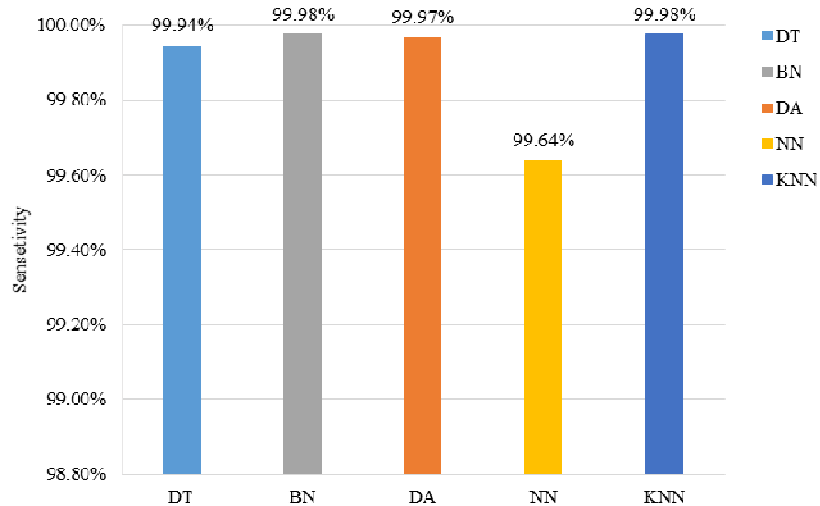


Fig. 11: Average sensitivity value for COIL-100 dataset using different data mining techniques

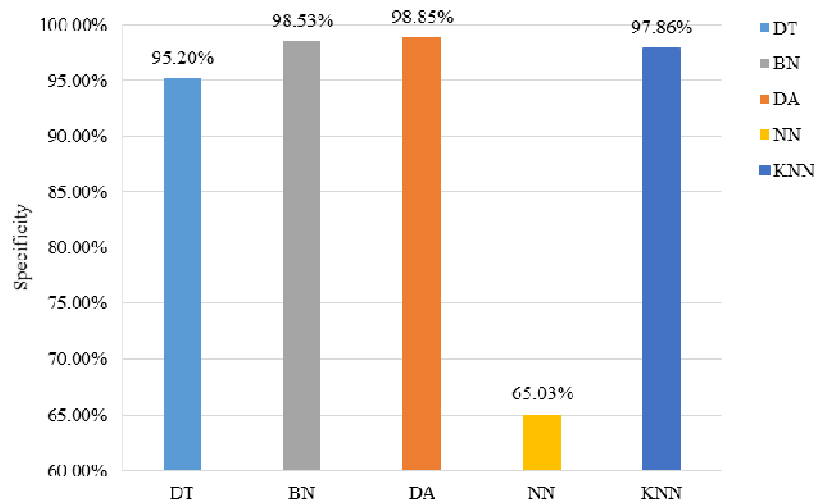


Fig. 12: Average specificity value for COIL-100 dataset using different data mining techniques

On the other hand, Fig. 7 to 9 present the average accuracy, sensitivity and specificity, respectively, which describe the overall system performance when applying the different classifiers on the Corel-1000 dataset. From the figures, we can observe that BN and DA classifiers outperform the other data mining techniques, regarding the accuracy, sensitivity and specificity. The average accuracy and specificity of the BN and DA classifiers are 99.9%. While the average sensitivity of the BN and DA classifiers are 99.5% and 99.4%, respectively.

Alternatively, Fig. 10 to 12 portray the average accuracy, sensitivity and specificity, respectively, for the COIL-100 dataset. The BN, DA and KNN classifiers show the best accuracy ( $\approx 100\%$ ) and sensitivity ( $\approx 100\%$ ), while DT classifier exhibits near optimal accuracy and sensitivity (with an insignificant difference). For the specificity, the best value (98.9%) is

attained by the DA classifier, whereas it is near optimal (98.5%) using BN classifier.

## Discussion

This paper studied the performance of five popular data mining techniques, NN, DT, BN, DA and KNN, together with BoVW model, for image classification. The comparisons delivered in this article provide a glance into how existing data mining classification techniques handle images with intrinsically different characteristics. Besides, how the BoVW representation using texture features (SURF and MSER descriptors) and color features (color correlograms and ICCV) collaborates with the data mining techniques for image classification. Two radically different datasets, including Corel-1000 and COIL-100, are used in our evaluation.

Encouragingly, all data mining techniques showed adequate ability to classify the considered images. Performance with different kind of images was found to vary by implemented data mining techniques. While DA and BN maintained similar performance with varying types of images, DT and KNN showed an observed increasing trend in the accuracy, sensitivity and specificity when classifying highly correlated images with large intra-class variation. NN technique, on the other hand, exhibited a significant improvement in accuracy and sensitivity with these highly correlated images, while the specificity showed an opposite trend.

The results of the Corel-1000 dataset indicate that DA and BN have achieved the best classification performance; accuracy, sensitivity and specificity are approximately 100%. While KNN has the lowest accuracy (94.6%), sensitivity (80.7%) and specificity (97.1%). Both DA and BN build a feature independent models (predictive model for DA and probabilistic model for BN) that may account for their superior performance. Whereas DA relies on a discriminative function that combines the extracted attributes, while BN counts on the degree of correlation of the obtained attributes. Unlike DA and BN, KNN does not build a classifier model and assigns equal weights to all attributes in order to compute the similarities between images, which may lead to classification errors due to the variation of the attributes significant effects within each image category. It is clear, also, that DT outperforms the NN. This is due to the powerful boosting method offered by C5.0 to increase the classification accuracy as well as its capability of removing unconstructive features. Whereas NN has many unknown parameters and it does not provide information covering the relative significance of these different parameters. Accordingly, it does not guarantee to attain more precise outcomes.

On the other hand, the results obtained from using COIL-100 dataset illustrate that DT, BN, DA and KNN attained the best accuracy and sensitivity ( $\approx 100\%$ ), while the best specificity ( $\approx 99\%$ ) is realized using BN and DA classifiers. Further, NN classifier has the lowest classification performance. It is obvious that the DT and KNN have a better performance for the COIL-100 dataset than that for the Corel-1000 dataset. This is because the former contains a uniform background that is highly uncorrelated with the foreground objects. Thus, it is less challenging than the real-world imagery of the Corel-1000 dataset that may contain contradictory cases, members of different classes, with indistinguishable attributes.

A factor contributing to the classification performance of the different implemented data mining techniques is the carefully derived features used. In this article, SURF and MSER descriptors are utilized to signify the texture features. While, color correlograms

and ICCV are used to denote color features. Combining texture and color descriptors produce robust, accurate and precise features. Furthermore, these features are used to build a BoVW model, which proved to be very efficient and robust in characterizing various kinds of images to facilitate classification. It showed an effective representation for natural seen images as well as highly correlated images that contains a single object.

Established along the empirical results, we can conclude that choosing an appropriate classifier to be applied for any particular image types is a heavy application dependent and thus a few of different classifiers may need to be examined before the optimal solution can be set up. Additionally, the BoVW that combines texture and color features provides a significant attributive for various kinds of images and collaborates properly with various data mining techniques to perform image classification task efficiently.

This comparative study reveals the general properties of the existing data mining techniques for image classification and provides a new vision into the strengths and shortcomings of these methods.

## Conclusion

This article discussed the performance of various data mining techniques for image classification using BoVW. SURF, MSER, color correlograms and ICCV features are extracted and used to build the dictionary of visual words. Thus, an image can be characterized as a feature vector that presents the number of occurrences of each visual word in that image. The obtained training images feature vectors are, finally, fed to the classification stage to obtain the query image category.

In the classification stage, five different kinds of data mining techniques including DT, BN, DA, NN and KNN are investigated and their classification performance is compared. Results on challenging image datasets have revealed that remarkably high level of classification performance is achieved using data mining methods. This has been shown to be true for object images that are clutter-free and relatively statistically uniform, as well as real-world images that contain considerable clutter and intra-class diversity.

Furthermore, our findings indicate that the BN and DA classification algorithms outperform the majority of the data mining algorithms in classification accuracy, sensitivity and specificity for both utilized datasets. Further, our evaluation highlights the influence of the utilized image types on the classification performance. The experiments show also that the representation of the images in the problem domain plays an important role in the performance of the application of different data mining techniques to this domain. Thus, the nature of the datasets and how the images are described affects the results. Therefore, it can be concluded that choosing a suitable data

mining technique for classifying images depends on the type, nature and representation of these images.

Future research may focus on combining several data mining methods for image classification as well as applying it on more diversity datasets.

## Acknowledgement

This work has been achieved at Computers and Systems Department, Electronics Research Institute.

## Author's Contributions

**Sahar Abd El Hameed Mokhtar:** Prepared the extracted features and applied different data mining techniques with suitable tuning and refining. Participated in implementing all experiments and analysing the experimental results. Participated in the research plan, the study organization and the manuscript writing.

**Heba Ahmed Elnemr:** Proposed the main idea of the image classification based on BoVW. Participated in implementing all experiments and analysing the experimental results. Participated in the research plan, the study organization and the manuscript writing.

## Ethics

This article is original and contains unpublished material. The corresponding author confirms that no ethical issues involved.

## References

- Abe, B.T., O.O. Olugbara and T. Marwala, 2012. Hyperspectral image classification using random forests and neural networks. Proceedings of the World Congress on Engineering and Computer Science, Oct. 24-26, San Francisco, USA.
- Acosta-Mendoza, N., A. Morales-González, A. Gago-Alonso, E.B. García-Reyes and J.E. Medina-Pagola, 2012. Image Classification Using Frequent Approximate Subgraphs. In: Progress in Pattern Recognition, Image Analysis, Computer Vision and Applications, Alvarez, L., M. Mejail, L. Gomez and J. Jacobo (Eds.), Springer, Berlin, Heidelberg, ISBN-10: 978-3-642-33275-3, pp: 292-299.
- Austin, P.C., J.V. Tu, J.E. Ho, D. Levy and D.S. Lee, 2013. Using methods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes. *J. Clin. Epidemiol.*, 66: 398-407. DOI: 10.1016/j.jclinepi.2012.11.008
- Bay, H., A. Ess, T. Tuytelaars and L.V. Gool, 2006. SURF: Speeded Up Robust Features. In: Computer Vision - ECCV, Leonardis, A., H. Bischof and A. Pinz (Eds.), Springer, Berlin, Heidelberg, ISBN-10: 978-3-540-33833-8, pp: 346-359.
- Benkrama, S., L. Zaoui and C. Charrier, 2013. Accurate image search using local descriptors into a compact image representation. *Int. J. Comput. Sci. Issues*, 10: 220-225.
- Chen, X., X. Gu and X. Hua, 2007. An improved color coherence vector method for CBIR. Proceedings of the Graduate Students Symposium of Communication and Information Technology Conference, (ITC'07), Beijing.
- Dash, S.K. and M. Panda, 2016. Image classification using data mining techniques. *Adv. Comput. Sci. Inform. Technol.*, 3: 157-162.
- Diz, J., G. Marreiros and A. Freitas, 2016. Applying data mining techniques to improve breast cancer diagnosis. *J. Med. Syst.*, 40: 203. DOI: 10.1007/s10916-016-0561-y
- Elnemr, H.A., 2016. Combining SURF and msr along with color features for image retrieval system based on bag of visual words. *J. Comput. Sci.*, 12: 213-222. DOI: 10.3844/jcssp.2016.213.222
- Fakhr, M. and A. Elsayad. 2012. Steel plates faults diagnosis with data mining models. *J. Comput. Sci.*, 8: 506-514. DOI: 10.3844/jcssp.2012.506.514
- Fernandez, G.C., 2002. Discriminant analysis, a powerful classification technique in data mining. Proceedings of the SAS Users International Conference, Apr. 14-17, Orlando, Florida, pp: 247-256.
- Han, J. and M. Kamber, 2006. Data Mining: Concepts and Techniques. 2nd Edn., Elsevier, Science and Technology, USA, ISBN-13: 978-1-55860-901-3, pp: 800.
- Hosseini, L. and R.S. Kandovan, 2017. Hyperspectral image classification based on hierarchical SVM algorithm for improving overall accuracy. *Adv. Remote Sens.*, 6: 66-75. DOI: 10.4236/ars.2017.61005
- Huang, J., S.R. Kumar, M. Mitra, W. Zhu and R. Zabih, 1997. Image indexing using color correlograms. Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 17-19, IEEE Xplore Press, San Juan, pp: 762-768. DOI: 10.1109/CVPR.1997.609412
- Huang, X., G. Zhao, W. Zheng and M. Pietikainen, 2012. Spatiotemporal local monogenic binary patterns for facial expression recognition. *Signal Process. Lett.*, 19: 243-246. DOI: 10.1109/LSP.2012.2188890
- Huang, X., G. Zhao, X. Hong, M. Pietikainen and W. Zheng, 2013. Texture Description with Completed Local Quantized Patterns. In: Image Analysis. Kämäräinen, J.K. and M. Koskela (Eds.), Springer, Berlin, Heidelberg, ISBN-10: 978-3-642-38886-6, pp: 1-10.
- Huang, X., S. Wang, X. Liu, G. Zhao and X. Feng *et al.*, 2016. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neuro. Comput.*, 175: 564-578. DOI: 10.1016/j.neucom.2015.10.096

- IBM SPSS, 2015. Modeler 17 modeling nodes. IBM Corporation.
- Jain, M. and P.S. Tomar, 2013. Review of image classification methods and techniques. *Int. J. Eng. Res. Technol.*, 2: 852-858.
- Kamavisdar, P., S. Saluja and S. Agrawal, 2013. A survey on image classification approaches and techniques. *Int. J. Adv. Res. Comput. Commun. Eng.*, 2: 1005-1009.
- Krig, S., 2014. Interest Point Detector and Feature Descriptor Survey. In: *Computer Vision Metrics*, Krig, S. (Ed.), Apress, Berkeley, CA, ISBN-10: 978-1-4302-5930-5, pp: 217-282.
- Li, T., S. Zhu and M. Ogihara, 2006. Using discriminant analysis for multi-class classification: An experimental investigation. *Knowl. Inform. Syst.*, 10: 453-472. DOI: 10.1007/s10115-006-0013-y
- Liu, D., S. Wang and F. Porikli, 2016. Spatial encoding of visual words for image classification. *J. Electronic Imag.*, 25: 1-6. DOI: 10.1117/1.JEL.25.3.033008
- Lowe, D.G., 2004. Distinctive image features from scale invariant keypoints. *Int. J. Comput. Vis.*, 60: 91-110. DOI: 10.1023/B:VISI.0000029664.99615.94
- Lu, D. and Q. Weng, 2007. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.*, 28: 823-870. DOI: 10.1080/01431160600746456.
- Luo, M. and K. Zhang, 2014. A hybrid approach combining extreme learning machine and sparse representation for image classification. *Eng. Applic. Artificial Intell.*, 27: 228-235. DOI: 10.1016/j.engappai.2013.05.012
- Matas, J., O. Chum, M. Urban and T. Pajdla, 2002. Robust wide baseline stereo from maximally stable extremal regions. *Proceedings of the British Machine Vision Conference, (MVC'02)*, BMVA Press, Cardiff, Wales, pp: 384-393. DOI: 10.5244/C.16.36
- Mitchell, T., 1997. *Machine Learning*. 1st Edn., McGraw-Hill, New York, ISBN-10: 0071154671, pp: 414.
- Murty, M.N. and V.S. Devi, 2011. Bayes Classifier. In: *Pattern Recognition*, Murty, M.N. and V.S. Devi (Eds.), Springer, London, ISBN-10: 978-0-85729-495-1, pp: 86-102.
- Naswale, P.P. and P.E. Ajmire, 2016. Image classification techniques-a survey. *Int. J. Emerg. Trends Technol. Comput. Sci.*, 5: 236-239.
- Nene, S., S. Nayar and H. Murase. 1996. Columbia Object Image Library (COIL-100). Technical Report CUCS-006-96, Department of Computer Science, Columbia University, New York.
- Ramayah, T., N.H. Ahmad, H.A. Halim, S.R.M. Zainal and M.C. Lo, 2010. Discriminant analysis: An illustrated example. *Afr. J. Bus. Manage.*, 4: 1654-1667.
- Singh, B.K., K. Verma and A.S. Thoke, 2016. Fuzzy cluster based neural network classifier for classifying breast tumours in ultrasound images. *Expert Syst. Applic.*, 66: 114-123. DOI: 10.1016/j.eswa.2016.09.006
- Spangler, W.E., J.H. May and L.G. Vargas, 1999. Choosing data-mining methods for multiple classification: Representational and performance measurement implications for decision support. *J. Manage. Inform. Syst.*, 16: 37-62. DOI: 10.1080/07421222.1999.11518233
- Tan, P.N., M. Steinbach and V. Kumar, 2004. *Introduction to Data Mining*. 1st Edn., Addison-Wesley Companion, ISBN-10: 0321321367.
- Wang, J., J. Li and G. Wiederhold, 2001. SIMPLiCity: Semantics-sensitive integrated matching for picture libraries. *IEEE Trans. Patt. Anal. Mach. Intell.*, 23: 947-963. DOI: 10.1109/34.955109
- Wang, L., K. Zhang, X. Liu, E. Long and J. Jiang *et al.*, 2017. Comparative analysis of image classification methods for automatic diagnosis of ophthalmic images. *Sci. Report*, 7: 1-11. DOI: 10.1038/srep41545
- Wang, R., K. Ding, J. Yang and L. Xue, 2016. A novel method for image classification based on bag of visual words. *J. Visual Commun. Image Representat.*, 40: 24-33. DOI: 10.1016/j.jvcir.2016.05.022
- Witten, I.H., E. Frank and M.A. Hall, 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd Edn., Elsevier Inc, ISBN-10: 0-12-374856-9, pp: 664.
- Yang, J., Y.G. Jiang, A.G. Hauptmann and C.W. Ngo, 2007. Evaluating bag-of-visual-words representations in scene classification. *Proceedings of the International Workshop on Multimedia Information Retrieval*, Sep. 24-29, ACM, Augsburg, Bavaria, Germany, pp: 197-206. DOI: 10.1145/1290082.1290111
- Zhang, G.P., 2000. Neural networks for classification: A survey. *IEEE Trans. Syst. Man Cybernet. Part C: Applic. Rev.*, 30: 451-462. DOI: 10.1109/5326.897072