Journal of Engineering and Applied Sciences 14 (20): 7627-7635, 2019
ISSN: 1816-949X


# A Survey Automatic Image Annotation Based on Machine Learning Models

[1,2]Myasar Mundher Adnan, [1,3]Mohd Shafry Mohd Rahim,
[2]Siraj Muneer Khaleel and [2]Karrar Al-Jawaheri
[1]Faculty of Engineering, School of Computing, University Technology of Malaysia,
Johar Bahru, Malaysia
[2]Islamic University, Najaf, Iraq
[3]UTM-IRDA Digital Media Institute of Human-Centred Engineering,
University Technology of Malaysia, Johar Bahru, Malaysia
maiser.monther@yahoo.com

**Abstract:** Image annotation has recently received much attention as a result of the rapid growth in image data. Several works have been proposed on AIA, especially, in the probabilistic modeling and classification-based methods. This study presents a review of the image annotation methods which has been published in the last 20 years. Emphasis is mainly on the machine learning models and the classification of the AIA methods into 5 categories of decision tree-based, Support Vector Machine (SVM)-based, k-Nearest Neighbor (kNN)-based, Deep Neural Network (DNN)-based and Bayesian-based AIAs. A comparison of the five types of AIA approaches was presented based on the underlying idea, feature extraction method, annotation accuracy, computational complexity and datasets. Furthermore, a review and explanation of the evaluation metrics used were presented. Emphasis was also placed on the to carefully consider these aspects during the development of new techniques and datasets for future image annotation tasks.

**Key words:** Image annotation, AIA, machine learning, image retrieval, development of new techniques, emphasis

## INTRODUCTION

The rapid advancement in multimedia and network has resulted in the generation of a huge volume of data which are stored or uploaded to the internet for browsing and sharing purposes. For an organized browsing in a large-scale image database there is a need for an efficient image retrieval technology which will shorten the required time for image retrieval. Image annotation is a key technology in the area of image retrieval and has become important that it has attracted a wide public interest. Image annotation implies the labeling of images with important keywords which will facilitate their retrieval and recognition. Additionally, image annotation can be classified into manual, automatic and semi-automatic image annotation. The manual annotation method is the conventional method which is very tedious as it requires a person to annotate each of the images that requires annotation. For the automatic method it is executed by a system with no human intervention. The semi-automatic method is the new method which is often applied in image retrieval and classification (Jeon *et al*., 2003; Chong *et al*., 2009). Numerous AIA methods have been developed and they have recorded various levels of efficiency. The major concept of the AIA is to accurately map the low-level features of an image such as the texture or color to some high-level semantic labels such as architecture, animals and landscape. There are several AIA methods and as such several AIA classification schemes exist such as learning-based, probability-based, non-probability-based, retrieval-based, unsupervised, supervised and semi-supervised methods (Cheng *et al*., 2018). This review focused on AIA classification based on supervised Machine Learning (ML). The ML supervised AIA classification methods are categorized into DT-based, SVM-based, kNN-based, DNN-based and Bayesian-based methods.

**Image segmentation and features extraction:** Performance in AIA depends on the features of the image and on the image segmentation. The major concept of the AIA is to accurately map the low-level features of an image such as the texture or color to some high-level semantic labels such as architecture, animals and landscape. The first in

**Corresponding Author:** Myasar Mundher Adnan, Faculty of Engineering, School of Computing,
University Technology of Malaysia, Johar Bahru, Malaysia, maiser.monther@yahoo.com

image segmentation is usually to extract the regions based on the image representation. The image is first divided by the segmentation algorithm into different components based on the similarity of the features (Zhang *et al*., 2012). Image segmentation mainly aims at generating a curve around an image. The evolution is terminated as soon as the curve coincides with the object's boundary. Being a difficult task many AIA techniques have used grid-based approaches to simplify this task by roughly segmenting the images into blocks (Wang *et al*., 2006; Jiang *et al*., 2009; Wan, 2011; Han and Qi, 2005; Xuelong *et al*., 2007; Gao *et al*., 2010; Alham *et al*., 2011; Wei *et al*., 2013; Fu, 2014; Shi and Malik, 2000; Zhu and Yuille, 1996; Tian, 2015; Wei *et al*., 2013) before extracting the visual features from the blocks. There is little computation in the block-based approach, however, it does not provide a full description of the semantic image components. Most of the current works (Mori *et al*., 1999; Figueiredo *et al*., 2001; Cusano *et al*., 2003 and Kwasnicka and Paradowski, 2006) have suggested solutions to image segmentation problem but have not satisfied the required aspirations. The features of an image are usually extracted for subsequent multimedia processing like tagging, recognition and classification. The intrinsic and discriminative contents of an image should be represented by the selected features from the original image and to achieve this some studies (Maree *et al*., 2005; Wang *et al*., 2006; Qi and Han, 2007; Lim *et al*., 2003) which are based on feature extraction have shown that the descriptors of the rectangular image regions are not strong enough to differentiate image transformations. However, some other experiments have proved their acceptability and can be automatically produced. This will be discussed in the later sections.

## MATERIALS AND METHODS

**Why using machine learning?:** The ML methods have been successfully used in different aspects of computer science and one of such aspect is in image analysis or automatic image annotation to be specific. The goal of ML is to construct computer systems that can learn and adapt from their experience (Dietterich, 2000).

Thus, the system becomes competent of the independent acquisition and integration of knowledge a process called machine learning. This learning results in a system that can progress with its own speed or performance of the process. The overall goal of machine learning is to improve efficiency and/or effectiveness of the system. The AIA has been an active area of research for several years now and has emerged from such research areas like cross-lingual machine translation and image recognition. AIA is possible today due to the increased computational, data transfer and data storage capabilities of the recent generation of computers. The AIA methods in existence during the past few years mainly depends on the ML approaches.

**Automatic image annotation methods using machine learning:** In this stuy, a brief review of methods for ML supervised AIA methods was presented.

**Decision Tree (DT) based AIA:** The DT is one of the commonest ML algorithms in use. It is a data mining induction framework which recursively divides data set using either a depth-first greedy or breadth-first approach. Hunt *et al*. (1966) compared the other classifier-learning methods (including those that are DT-based) in terms of their industrial and commercial applications (Michie *et al*., 1994). Jiang *et al*. (2009) proposed image annotation using DT-based Bayesian (DTB) ML. The DTB is an improved DT framework which comprised of a DT classifier and an improved NB classifier. It can be deployed in image training and used to automatically annotate an image in the unknown image database using the established DTB principles. The DTB has been previously shown to be accurate in most classification problems and also useful in obtaining an appropriate set of rules from a huge number of instances. Wan (2011) presented a Simple DT (SDT) classification algorithm as an improved version of the DT algorithm. The SDT calculate model by the heuristic search of model space for fast decision tree algorithm. Additionally Wan (2011) apply the SDT algorithm in an image annotation system for the faster and efficient classification of a large number of training data. The SDT algorithm is based on the corel image library training set which could perform both image labeling and classification, although, the experimental finding was not satisfactory. Li *et al*. (2015) proposed an AIA method which was based on fuzzy association rule and DT. They proposed the integration of Fuzzy Association Rules (FARs) and DT for the provision of a way of obtaining more optimal annotation rules. There are several advantages of the proposed method on one hand, the optimal FARs which determines the level of image relationship between concepts and features was obtained. Patil and Kolhe (2014) discussed the use of AIA based on DT and rough sets techniques. A classifier was proposed based on the use of Rough Set (RS) for the classification of images from corel image dataset for annotation purposes. The feature-texture performance of the classifier was poor compared to four other feature combinations. The DT cannot differentiate images, since, IF-ELSE is generated by the raw features based on the binary tree for

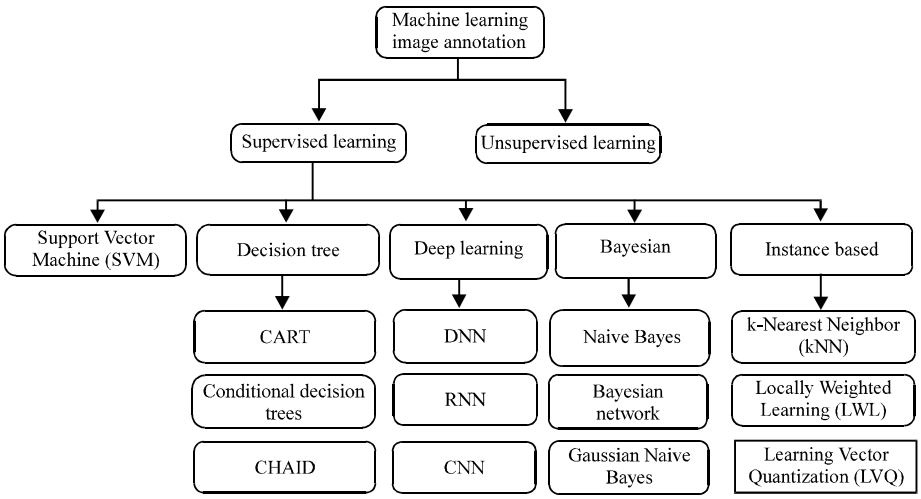


Fig. 1: Machine learning based on AIA

image classification. Comparatively, rough sets provide better performances, since, the approximation boundaries play a role during rule construction. The cut set is generated on the original data to discretize the actual values from the dataset (Fig. 1).

**Support Vector Machine (SVM) based on AIA:** The SVM is another ML algorithm which is based on the statistical learning theory. It works based on the Structural Risk Minimization (SRM) principle instead of the empirical risk minimization of huge samples. It is a supervised learning model which is based on learning algorithms theory with the major idea of adjusting a discriminating function to the level of making an optimal use of the separability information of boundary cases. Hatem *et al.* (2016) introduced an automatic sports image annotation framework which was based on a sequential step comprised of image segmentation, extraction of features, image annotation and classification of image segments. Experiments with the SVM and DT classifiers proved Cusano *et al.* (2003) presented an automatic digital photographs annotation scheme. The scheme can annotate digital photographs by assigning regions to 7 different classes (sky, skin, vegetation, snow, water, ground and buildings) before independently classifying each class using a multi-class SVM. It was observed that the misclassified pixels were either on the inter-class boundary or within very dark areas where there are an insufficient color and texture information. These errors suggest that the tool is not suitable to be used as a segmentation tool as a more specific segmentation method will be required for this purpose. Goh *et al.* (2005) proposed the annota-tion of images using one-class, two-class and multiclass SVMs in order to support the retrieval of image keywords. An accurate image mapping is required to achieve an automatic annotation. Han and Qi (2005) suggested a combination of MIL-based and global-feature-based SVMs for an efficient image annotation. In this system, the image is partitioned into several blocks instead of being segmented into homogeneous regions. The bag features of an image are obtained using the MIL and used as the input to the SVMs when finding the optimal hyperplanes for categorizing the training images. To solve the likely Rotation Scaling and Translation (RST) variant problems, the global-feature-based SVM whose inputs are the global color and edge histograms was developed for categorizing the training images. The incorporation of multiple SVMs for AIA was proposed based on the integration of 2 sets of SVMs, namely the MIL-based and the global-feature-based SVMs for image annotation (Qi and Han, 2007). The MIL was applied to extract the MIL-based bag features from the image blocks before applying the enhanced Diversity Density (DD) algorithm and another faster searching algorithm for the improvement of the accuracy and efficiency of the system. Bhargava *et al.* (2014) proposed a mechanism which uses the Speeded-Up Robust Features (SURF) technique the extraction of the important features before using SVM as a classifier for the training of the images and their classification into groups. This object selection phase helps in the partitioning of an image into different groups based on the objects present in the image. Therefore, there is no need to extract all the image features when a new image is encountered. Xuelong *et al.* (2007) utilized only SVM to achieve a semi-automatic image annotation. The image collection process was divided into 2 parts, 1 part of the manual annotation process and the other part

for testing after classification using SVM. Gao *et al*. (2010) proposed a hierarchical annotation method which considered that human visual identification of a scenery object is a rough tone hierarchical process. First, the input image is partitioned into regions before roughly labeling each segmented region with a generalized keyword using a multi-classification SVM. A prototype image annotation system was later developed whose preliminary findings showed it as an effective hierarchical annotation scheme. Alham *et al*. (2011) proposed and evaluated a MapReduce-based distributed SVM which depended on the parallelism, scalability and resiliency of the MapReduce framework for AIA. The MRSMO algorithm was evaluated for performance and the results showed the MRSMO to significantly reduce the training time while still maintaining a good level of accuracy in both multiclass and binary classifications. Wei *et al*. (2013) proposed an AIA approach which is based on a multi-class SVM with an ontology for achieving a higher level of accuracy. A chose semantic dictionary wordnet with hierarchy was derived from the text ontology for the calculation of the correlations between the keywords. Bags of visual words model were virtually used to present the visual feature of an image before applying a multi-class SVM with mixed kernel. The results of the evaluation showed the ability of the combined ontology and multi-class SVM classifier to provide a potential way of performing and improving AIA. Jin and Jin (2017) developed an Improved Localized Multiple Kernel Learning (ILMKL) algorithm using a weight function for describing an image to belong to a semantic class as well as a novel MIA approach based on ILMKL called MIAILMKL. Fu (2014) proposed an automatic image semantic annotation technology which is SVM-based. First, the image visual features were mapped into one or more rough concept images using annotation before pre-processing the web page text information. Finally, the high keyword similarity was selected as an image annotation. The experimental results showed the scheme to be effective in improving the precision of image tagging. Most times, even though the SVM can achieve a satisfactory annotation performance and presents to be easily implemented, it is still prone to several problems which require further investigation. Firstly, the SVM is susceptible to the class-imbalance problem, meaning that it performs poorly on imbalanced datasets. Secondly, automatic image segmentation is well known as an open problem in computer vision (Shi and Malik, 2000; Zhu and Yuille, 1996) and as such no system can achieve a perfect segmentation outcome. Additionally, SVM is computationally costly. Thirdly, the training time of SVM generally increases linearly with the number of training images because with each image, the time complexity increases. Fourthly, regarding large-scale image annotation it is believed that different categories are visually similar in the feature space and this may result in the overlapping of features and performance degradation. Hence, the integration of the contextual and correlation information of different annotations into the SVM image annotation process can help in the improvement of AIA performance (Tian, 2015).

**k-Nearest Neighbour (kNN) based AIA:** The kNN is one of the lazy learning or instance-based learning algorithms whose function is mainly localized and all computation with held until classification. The kNN is one of the simplest ML algorithms which is helpful for both regression and classification processes where it can help to weight the neighbour's contributions such that the nearer neighbours contribute higher to the average compared to the distant neighbours (Chu *et al*., 2014). Liu *et al*. (2012) proposed a graph-based dimensionality reduction approach for kNN-based image annotation through the adaptation of the graph-based Locality Sensitive Discriminant Analysis (LSDA) method of (Liu *et al*., 2012) in a multi-label setting. In the LSDA, the nearest neighbour graph is constructed before splitting the graph into within class and between class graphs based on the class labels. Finally, a linear low dimensional embedding is established by considering both graphs. To fit LSDA to multi-label learning tasks, two images relevant or irrelevant ought to be defined based on the label vectors relationship. Then, the relevant and irrelevant graphs are constructed instead of constructing the within class and between class graphs based on the relevant and irrelevant image features that are virtually similar. This method can pull the neighbourhood of relevant images closer and push the irrelevant images further away due to the pull-push strategy deployed. Then, the relevant and irrelevant graphs are specifically designed and this is proper for kNN-based image annotation. The method was proven to be efficient based on the experimental results. Bakliwal and Jawahar (2015) depended on the simplicity of the kNN classifier to reduce the required amount of training data during annotation tasks with a simultaneous improvement of the active learning framework performance. There are two factors that determine the performance of a typical annotation algorithm, feature representation and the underlying similarity metric and size and quality of the training dataset. There are many disadvantages of using a fixed length annotation method, especially, when the annotation length is either more or less than the mean annotation length. Chu *et al*. (2014) suggested an image annotation scheme based on the

combination of SVM and kNN algorithms. The framework combined three kinds of features (EDH, GLCM and area weighted HSV colour histogram) as the image content feature vector (Chu *et al*., 2014). Ji *et al*. (2017) proposed a new method called kNN-GSR. First, they obtained subsets of the training images based on their labels. Secondly, the used the traditional 2P kNN method to obtain the visual neighbours of the test image in each subset. Then, the found the semantic neighbours of the visual neighbours in each subset. Finally, according to Bayes theorem, the used all the neighbours to assign the importance for each label. The experimental results showed the effectiveness of the proposed method in comparison to the state of the art methods (Ji *et al*., 2017). The kNN-based image annotation method was proven successful, however, it suffers from two issues which are a high computational cost and the difficulty of finding semantically similar images. Nevertheless, due to their high time and space complexity, the graph-based learning methods are not realistic in the real world. The performance of these models become even worse when the vocabulary becomes large.

**Deep Neural Network (DNN) based AIA:** A DNN is a network characterized by a certain level of complexity it is a neural network consisting of more than two layers. The DNNs depends on sophisticated mathematical modeling for data processing in complex ways. Zhu *et al*. (2015) developed a new framework of multimodal deep learning network for learning intermediate representations and to provide a good network initialization. Then, the distance metric functions on each individual modality were optimized using back propagation, finally, an optimization of the combinational weights of different modalities was performed by applying the exponentiated gradient online learning algorithm. A further deep learning research which will focus on the determination of the number of feature dimensions for achieving a satisfactory system performance for a given neural network framework is necessary. Another aspect to consider is the mechanism to be used to enhance a given deep learning architecture and improve its robustness. Yang *et al*. (2015) proposed a new MVSAE Model for a joint establishment of the correlations between high level semantic keywords and low level image features for automatic image annotation. First, the SAE was modified by using an iteration algorithm and a sigmoid function predictor. Then, the image keywords were solved with imbalanced distribution. The influence of the imbalance learning method at different levels of keyword frequency varies. The F1 score decreases slightly towards high frequency keywords because of the tendency of a low frequency to cause a misclassification of a high frequency keyword. Contrarily, the low-level frequency key words present a better performance compared to the original SAE. A Multi-View Stacked Auto-Encoder (MVSAE) framework has been proposed by Yang *et al*. (2015) for finding the correlations between high level semantic information and low level visual features. Experiments No. 3 popular datasets proved the effectiveness of the proposed framework in achieving a favourable performance for image annotation. The DNNs with multiple nonlinear hidden layers can learn complex input output relationships, however, the network can be exposed to local optima and convergence difficulty due to the nonlinear mapping between the outputs and the inputs when using BP algorithm (Vincent *et al*., 2010).

**Bayesian based AIA:** Ivasic-Kos *et al*. (2014) proposed a multi-level image annotation models with 2 phases. The first phase involves the use of an NBC to classify the low level image features into classes while the second phase involves the use of a knowledge representation scheme based on fuzzy petri net for the expansion of the vocabulary level and to incorporate multi-level semantic concepts which are related to images into image annotations. Rui *et al*. (2005) and Heller and Ghahramani (2006) proposed a semi-NBC approach through the incorporation of clustering with pair wise constraints for AIA. Experiments have shown the method to considerably improve the annotation performance. Darwish *et al*. (2016) propose a novel effort towards a Multi-Instance Multi-Label image annotation (MIML). Here, images are first segmented using the Otsu method which selects an optimum threshold through the maximization of the image's intracluster variance. The Otsu method is modified using FA in order to optimize the runtime and the segmentation accuracy. Simonyan and Zisserman (2014) presented a Bayesian framework for content based image retrieval. The advantage of this method is that several images are used to perform retrieval instead using just a single image. The method achieved good results based on the Bayesian criterion and based on the marginal likelihood of finding the images that are most likely to belong to a query group. From the evaluations presented, so far, it is evident that the semi-Naive Bayes is more effective compared to the NB. Note that a Bayesian learning model is deployed when building upon regions that are obtained through latent topic allocation. Meanwhile, this is a very complex model compared to other ML models. With the involvement of several conditional probabilities, the reliability of the system may not be easily ascertained.

**Summaries and concludes:** The performance of various DT and SVM methods on the corel-5K dataset is presented in Table 1. It could be observed from the table

Table 1: Comparison of image annotation based on decision tree, rough set and SVM models on corel-5K datasets

| Sources | Classifiers | Accuracy | Image datasets |
|---|---|---|---|
| Cusano *et al.* (2003) | Multi-class SVM | 89.71 | Corel |
| Goh *et al.* (2005) | One class-two clss SVM | 61.50 | 25K-image |
| Han and Qi (2005) | SVMS | 81.40 | Corel |
| Qi and Han (2007) | Multiple SVM | 88.80 | Corel |
| Xuelong *et al.* (2007) | SVM | 83.50 | Corel |
| Jin and Jin (2017) | SVM | 66.60 | Corel stock |
| Nasullah Khalid *et al.* | MRSMO, SVM | 93 | Corel |
| Wei *et al.* (2013) | Ontology multi-class SVM | Low | Corel5K |
| Fu (2014) | SVM | 78.33 | Corel 2000 |
| Chu *et al.* (2014) | DTB | 81.85 | Corel |
| Wan (2011) | STD | 79.80 | Corel |
| Patil and Kolhe (2014) | DT and RS | 57.40 | Corel |

Table 2: Performance comparison of image annotation based on kNN, DNN, Bayes and semi Bayes models on corel 5K datasets

| Methods | P | R | F1 | N+ | Datasets |
|---|---|---|---|---|---|
| DNN (72) | 34.00 | 31.00 | 33.00 | | Corel 5K |
| DNN+softmax regression (73) | 0.37 | 0.34 | 0.36 | | Corel 5K |
| MVSAE (66) | 37.00 | 47.00 | 42.00 | 175 | Corel 5K |
| KREPN (68) | 34.00 | 26.00 | 9.00 | | Coreldataset |
| SNB (70) | 36.00 | 45.00 | 41.00 | | Corel 5K |
| GS (31) | 30.00 | 33.00 | 32.00 | 146 | Corel 5K |

that DTB (Jiang *et al*., 2009) and STD (Wan, 2011) recorded the best performances among of DT models. Additionally, the performance of the SVM methods on the corel dataset was compared, since, a large amount of data is required by the SVM to achieve a proper training. The comparison result showed that MRSMO and SVM (Alham *et al*., 2011) generally, achieved the best performance while the least performance was recorded by the one class-two class SVM (Goh *et al*., 2005) and ontology multi-class SVM (Wei *et al*., 2013).

The methods were compared in terms of their performances using several image annotation effectiveness evaluation measures such as Precision (P), Recall (R), measure (the weighted harmonic mean of precision and recall), N+ (the number of tags that has non-zero recall) and Mean Average Precision (MAP). Table 2 showed the outcome of the performance comparison of the ML models on the corel dataset. From the results, MVSAR (Yang *et al*., 2015) outperformed the other methods.

## Datasets

**Corel-5K (Verma and Jawahar, 2012):** This dataset consists of 4500 training images and 499 testing images with each image annotated with up to 5 labels (approximately 3.4 labels per image). The corel-5K is one of the oldest datasets for image annotation.

**ESP game (Duygulu *et al*., 2002):** The ESP game consists of 18689 training images and 2081 testing images and each image is annotated with up to 15 labels (approximately 4.7 labels per image). The dataset was formed from an online game where 2 players are meant to assign labels to a given image with a point scored for each common label. As such, several participants are involved in the manual annotation task thereby making it a challenging dataset.

**IAPR TC-12 (Von Ahn and Dabbish, 2004):** This dataset consists of 17665 training images and 1962 testing images with each image annotated with up to 23 labels (approximately 5.7 labels per image). Each image has a long description in several languages. Heller and Ghahramani (2006) and Simonyan and Zisserman (2014) used the English language to extract nouns from the image descriptions and considered them as annotations. Since, then it has been a widely used method for the evaluation of image annotation methods.

**NUS-WIDE:** The NUS-WIDE is the largest publicly available image annotation dataset. It is comprised of 269648 images which were downloaded from flickr and with 81 labels in the vocabulary. Each image in the dataset is annotated with up to 3 labels (approximately 2.40 labels per image). Based on earlier reports (Wang *et al*., 2006; Zhu *et al*., 2015) images with labels were discarded in this report, leaving a net 209,347 images which were split into 125000 training images and 80000 images for testing using the split method originally proposed by the researchers of the dataset.

**MS-COCO (Chua *et al*., 2009):** The MS-COCO is next to NUS-WIDE in size and a popular dataset for image annotation. It is comprised of 82783 training images with 80 labels with each image being annotated with an average of 2.9 labels. Although, it is not available publicly, it is used for image recognition.

## RESULTS AND DISCUSSION

**Evaluation metrics:** Several metrics can be used for the evaluation of the performance of several AIA methods. Some of the evaluation measures are average precision per label, average F1-score per label, average recall per label and the normalized N+ score (Rakhmadi *et al*., 2010).

Recall and precision for a given keyword, let M1 represents the number of images contained in the test dataset which has been annotated with the label, M2 to represent the number of images which has been correctly annotated with the labels and M3 to represent the number of images which has been annotated with the label using the ground-truth data. Then, recall may be expressed as M2/M3 and precision as M2/M1. The mean precision and recall for the data set are achieved as the average of the precision and recall values over Y. An analysis of the balance between precision and recall requires the calculation of the mean F1-score as F1 = 2PR/(P+R). N+: N+ represents the number of keywords which have been correctly assigned to at least one test image. It also represents the number of keywords with the positive recall. A high N+ value indicates a good performance of the corresponding AIA method (Ruzinoor *et al*., 2012).

**Performance comparison:** The models were evaluated and compared based on their performance on the corel 5K (Rad *et al*., 2016). Their performances were evaluated using the previously mentioned evaluation metrics and their observed performances are presented in Table 1 and 2. We compared the performance of various ML methods as shown in Table 1 and 2 and from the results, almost all the ML-based models suffered from the problem of a semantic gap between high-level concepts and low-level image features. A great problem of image annotation is a proper features extraction. Several conclusions can be generally drawn as follows, firstly, there is a semantic gap between high-level semantics and low-level image features, secondly, class imbalance is a major problem in multilabel image annotation where the frequently occurring labels suppress the contribution of the rarely occurring labels, thirdly, image annotation systems often achieve a low recall and a high precision. The major function of an image annotation system is to ensure a balance between recall and precision through the improvement of the recall scores without having much effect on the precision (Jabal *et al*., 2009). Finally, image annotation frameworks require a prolonged time and are computationally complex, especially, when the training dataset is large. The Bayesian model is more complicated compared to the other ML models and its reliability cannot be easily ascertained due to the involvement of several conditional probabilities. The DT is not capable of discriminating images, since, the raw features classify images by generating IF-ELSE-based binary trees (Rahim *et al*., 2017).

## CONCLUSION

A review of the AIA methods and their classification decision tree based-AIA, Support Vector Machine (SVM) based AIA, k-Nearest Neighbor (kNN) based AIA, Deep Neural Network (DNN) based AIA and Bayesian-based AIA were presented in this study. A comparison of the five types of AIA approaches has been presented based on the underlying idea, the feature extraction method, annotation accuracy, computational complexity and datasets. Furthermore, a review and explanation of the evaluation metrics used to evaluate AIA methods were presented.

## REFERENCES

Alham, N.K., M. Li, Y. Liu and S. Hammoud, 2011. A mapreduce-based distributed SVM algorithm for automatic image annotation. Comput. Math. Appl., 6: 2801-2811.

Bakliwal, P. and C.V. Jawahar, 2015. Active learning based image annotation. Proceedings of the 2015 5th National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), December 16-19, 2015, IEEE, Patna, India, ISBN:978-1-4673-8564-0, pp: 1-4.

Bhargava, A., S. Shekhar and K.V. Arya, 2014. An object based image retrieval framework based on automatic image annotation. Proceedings of the 2014 9th International Conference on Industrial and Information Systems (ICIIS), December 15-17, 2014, IEEE, Gwalior, India, ISBN:978-1-4799-6499-4, pp: 1-6.

Cheng, Q., Q. Zhang, P. Fu, C. Tu and S. Li, 2018. A survey and analysis on automatic image annotation. Pattern Recognit., 79: 242-259.

Chong, W., D. Blei and F.F. Li, 2009. Simultaneous image classification and annotation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR 2009, June 20-25, 2009, IEEE, Miami, Florida, ISBN:978-1-4244-3992-8, pp: 1903-1910.

Chu, G., K. Niu and B. Tian, 2014. Automatic image annotation combining SVMS and KNN algorithm. Proceedings of the 2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems (CCIS), November 27-29, 2014, IEEE, Shenzhen, China, ISBN:978-1-4799-4720-1, pp: 13-17.

Chua, T.S., J. Tang, R. Hong, H. Li and Z. Luo et al., 2009. NUS-WIDE: A real-world web image database from National University of Singapore. Proceedings of the ACM International Conference on Image and Video Retrieval (CIVR '09), July 08-10, 2009, ACM, New York, USA., ISBN:978-1-60558-480-5, pp: 1-9.

Cusano, C., G. Ciocca and R. Schettini, 2003. Image annotation using SVM. Proceedings of the SPIE Conference on Internet Imaging Vol. 5304, December 15, 2003, SPIE, San Jose, California, USA., pp: 1-9.

Darwish, S.M., M.A. El-Iskandarani and G.M. Shawkat, 2016. Automatic multi-label image annotation system guided by firefly algorithm and bayesian classifier. Proceedings of the World Congress on Engineering (WCE 2016) Vol. 1, Jun 29-July 1, 2016, IAENG, Hong Kong, China, ISBN:978-988-14048-0-0, Page: 1-8.

Dietterich, T.G., 2000. Hierarchical reinforcement learning with the MAXQ value function decomposition. J. Artif. Intell. Res., 13: 227-303.

Duygulu, P., K. Barnard, J.F.D. Freitas and D.A. Forsyth, 2002. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. Proceedings of the European Conference on Computer Vision, May 28-31, 2002, Springer, Copenhagen, Denmark, ISBN:978-3-540-43748-2, pp: 97-112.

Figueiredo, M.A.T., A. Vailaya, A.K. Jain and H.J. Zhang, 2001. Image classification for content-based indexing. IEEE Trans. Image Processing, 10: 117-130.

Fu, J.C., 2014. An approach of image semantic automatic tagging based on SVM. Appl. Mech. Mater., 530: 382-385.

Gao, Y.Y., Y.X. Yin and T. Uozumi, 2010. A Hierarchical image annotation method based on SVM and semi-supervised EM. Acta Automat. Sinica, 36: 960-967.

Goh, K.S., E.Y. Chang and B. Li, 2005. Using one-class and two-class SVMs for multiclass image annotation. IEEE. Trans. Knowl. Data Eng., 17: 1333-1346.

Han, Y. and X. Qi, 2005. A complementary SVMs-based image annotation system. Proceedings of the 2005 IEEE International Conference on Image Processing, September 14, 2005, IEEE, Genova, Italy, ISBN:0-7803-9134-9, pp: I-1185.

Hatem, Y., S. Rady, R. Ismail and K. Bahnasy, 2016. Automatic content description and annotation of sport images using classification techniques. Proceedings of the 10th International Conference on Informatics and Systems (INFOS '16), ACM, New York, USA, ISBN:978-1-4503-4062-5, pp: 88 94-10.1145/2908446.2908458.

Heller, K.A. and Z. Ghahramani, 2006. A simple Bayesian framework for content-based image retrieval. Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) Vol. 2, June 17-22, 2006, IEEE, New York, USA, ISBN:0-7695-2597-0, pp: 2110-2117.

Hunt, E., J. Martin and P. Stone, 1966. Experiments in Induction. Academic Press, New York.

Ivasic-Kos, M., I. Ipsic and S. Ribaric, 2014. Multi-level image annotation using bayes classifier and fuzzy knowledge representation scheme. WSEAS. Trans. Comput., 13: 635-644.

Jabal, M.F.A., M.S.M. Rahim, N.Z.S. Othman and Z. Jupri, 2009. A comparative study on extraction and recognition method of CAD data from cad drawings. Proceedings of the 2009 International Conference on Information Management and Engineering, April 3-5 2009, IEEE, Kuala Lumpur, Malaysia, pp: 709-713.

Jeon, J., V. Lavrenko and R. Manmatha, 2003. Automatic image annotation and retrieval using cross-media relevance models. Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, July 28-August 01, 2003, ACM, Toronto, Canada, ISBN:1-58113-646-3, pp: 119-126.

Ji, Q., L. Zhang and Z. Li, 2017. KNN-based image annotation by collectively mining visual and semantic similarities. KSII. Trans. Internet Inf. Syst., 11: 4476-4490.

Jiang, L., J. Hou, Z. Chen and D. Zhang, 2009. Automatic image annotation based on decision tree machine learning. Proceedings of the 2009 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, October 10-11, 2009, IEEE, Zhangijajie, China, ISBN:978-1-4244-5218-7, pp: 170-175.

Jin, C. and S.W. Jin, 2017. A multi-label image annotation scheme based on improved SVM multiple kernel learning. Proceedings of the 8th International Conference on Graphic and Image Processing (ICGIP 2016) Vol. 10225, February 8, 2017, SPIE, Bellingham, Washington, pp: 1-6.

Kwasnicka, H. and M. Paradowski, 2006. Multiple class machine learning approach for an image auto-annotation problem. Proceedings of the 6th International Conference on Intelligent Systems Design and Applications Vol. 2, October 16-18, 2006, IEEE, Jinan, China, pp: 347-352.

Li, Z., L. Li, K. Yan and C. Zhang, 2015. Automatic image annotation based on fuzzy association rule and decision tree. Proceedings of the 7th International Conference on Internet Multimedia Computing and Service (ICIMCS '15), August 19-21, 2015, ACM, New York, USA, ISBN:978-1-4503-3528-7, pp: 1-6.

Lim, J.H., Q. Tian and P. Mulhem, 2003. Home photo content modeling for personalized event-based retrieval. IEEE. Multimedia, 10: 28-37.

Liu, X., R. Liu, F. Li and Q. Cao, 2012. Graph-based dimensionality reduction for KNN-based image annotation. Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), November 11-15, 2012, IEEE, Tsukuba, Japan, ISBN:978-1-4673-2216-4, pp: 1253-1256.

Maree, R., P. Geurts, J. Piater and L. Wehenkel, 2005. Random subwindows for robust image classification. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) Vol. 1, June 20-25, 2005, IEEE, San Diego, California, USA., ISBN:0-7695-2372-2, pp: 34-40.

Michie, D., D.J. Spiegelhalter and C.C. Taylor, 1994. Machine Learning, Neural and Statistical Classification. Ellis Horword, Chichester, England, ISBN:9780131063600, Pages: 289.

Mori, Y., H. Takahashi and R. Oka, 1999. Image-to-word transformation based on dividing and vector quantizing images with words. Proceedings of the 1st International Workshop on Multimedia Intelligent Storage and Retrieval Management, October 30, 1999, Orlando, Florida, pp: 1-9.

Patil, M.P. and S.R. Kolhe, 2014. Automatic image annotation using decision trees and rough sets. Intl. J. Comput. Sci. Appl., 11: 38-49.

Qi, X. and Y. Han, 2007. Incorporating multiple SVMs for automatic image annotation. Pattern Recognit., 40: 728-741.

Rad, A.E., M.S.M. Rahim, A. Rehman and T. Saba, 2016. Digital dental X-ray database for caries screening. 3D Res., Vol. 7. 10.1007/s13319-016-0096-5

Rahim, M.S.M., A. Norouzi, A. Rehman and T. Saba, 2017. The 3D bones segmentation based on CT images visualization. Intl. J. Med. Sci., 28: 3641-3644.

Rakhmadi, A., N.Z.S. Othman, A. Bade, M.S.M. Rahim and I.M. Amin, 2010. Connected component labeling using components neighbors-scan labeling approach. J. Comput. Sci., 6: 1099-1107.

Rui, S., W. Jin and T.S. Chua, 2005. A novel approach to auto image annotation based on pairwise constrained clustering and semi-naive Bayesian model. Proceedings of the 11th International Conference on Multimedia Modelling, January 12-14, 2005, IEEE, Melbourne, Australia, ISBN:0-7695-2164-9, pp: 322-327.

Ruzinoor, C.M., A.R.M. Shariff, B. Pradhan, M.R. Ahmad and M.S.M. Rahim, 2012. A review on 3D terrain visualization of GIS data: Techniques and software. Geo-Spatial Inform. Sci., 15: 105-115.

Shi, J. and J. Malik, 2000. Normalized cuts and image segmentation. IEEE. Trans. Pattern Anal. Mach. Intell., 22: 888-905.

Simonyan, K. and A. Zisserman, 2014. Very deep convolutional networks for large-scale image recognition. J. Comput. Vision Pattern Recognit., 1: 1-14.

Tian, D., 2015. Support vector machine for automatic image annotation. Intl. J. Hybrid Inf. Technol., 8: 435-446.

Verma, Y. and C.V. Jawahar, 2012. Image annotation using metric learning in semantic neighbourhoods. Proceedings of the European Conference on Computer Vision, October 7-13, 2012, Springer, Berlin, Heidelberg, Germany, ISBN:978-3-642-33711-6, pp: 836-849.

Vincent, P., H. Larochelle, I. Lajoie, Y. Bengio and P.A. Manzagol, 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res., 11: 3371-3408.

Von Ahn, L. and L. Dabbish, 2004. Labeling images with a computer game. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04), April 24-29, 2004, ACM, New York, USA., ISBN:1-58113-702-8, pp: 319-326.

Wan, S., 2011. Image annotation using the simple decision tree. Proceedings of the 2011 5th International Conference on Management of E-Commerce and E-Government, November 5-6, 2011, IEEE, Hubei, China, ISBN:978-1-4577-1659-1, pp: 141-146.

Wang, C., F. Jing, L. Zhang and H.J. Zhang, 2006. Image annotation refinement using random walk with restarts. Proceedings of the 14th ACM International Conference on Multimedia (MM '06), October 23-27, 2006, ACM, New York, USA, ISBN:1-59593-447-2, pp: 647-650.

Wei, Z., X. Luo and F. Zhou, 2013. Ontology based automatic image annotation using multi-class SVM. Proceedings of the 2013 7th International Conference on Image and Graphics (ICIG), July 26-28, 2013, IEEE, Qingdao, China, ISBN:978-0-7695-5050-3, pp: 434-438.

Xuelong, H., Z. Yuhui and Y. Li, 2007. A new method for semi-automatic image annotation. Proceedings of the 2007 8th International Conference on Electronic Measurement and Instruments, August 16-18, 2007, IEEE, Xi'an, China, ISBN:978-1-4244-1135-1, pp: 2-866.

Yang, Y., W. Zhang and Y. Xie, 2015. Image automatic annotation via multi-view deep representation. J. Visual Commun. Image Represent., 33: 368-377.

Zhang, D., M.M. Islam and G. Lu, 2012. A review on automatic image annotation techniques. Pattern Recogn., 45: 346-362.

Zhu, S., Z. Shi, C. Sun and S. Shen, 2015. Deep neural network based image annotation. Pattern Recognit. Lett., 65: 103-108.

Zhu, S.C. and A. Yuille, 1996. Region competition: Unifying snakes, region growing and Bayes/MDL for multiband image segmentation. IEEE. Trans. Pattern Anal. Mach. Intell., 18: 884-900.