Comparison of Handcrafted Features and Deep Learning in Classification of Medical X-ray Images

Mohammad Reza Zare
School of Information Technology
Monash University Malaysia
Subang Jaya, Malaysia
mohammad.reza@monash.edu

David Olayemi Alebiosu
School of Information Technology
Monash University Malaysia
Subang Jaya, Malaysia

Sheng Long lee
School of Information Technology
Monash University Malaysia
Subang Jaya, Malaysia

***Abstract*— The rapid growth and spread of radiographic equipment in medical centres have resulted in a corresponding increase in the number of medical X-ray images produced. Therefore, more efficient and effective image classification techniques are required. Three different techniques for automatic classification of medical X-ray images were compared. A bag-of-visual-words model and a Convolutional Neural Network (CNN) were used to extract features from the images. The two groups of extracted feature vectors were each used to train a linear support vector machine classifier. Third, a fine-tuned CNN was used for end-to-end classification. A pre-trained CNN was used to overcome dataset limitations. The three techniques were evaluated on the ImageCLEF 2007 medical database. The database provides medical X-ray images in 116 categories. The experimental results showed that fine-tuned CNN outperforms the other two techniques by achieving per class classification accuracy above 80% in 60 classes compared to 24 and 26 classes for bag-of-visual-words and CNN extracted features respectively. However, certain classes remain difficult to classify accurately such as classes in the same sub-body region due to inter-class similarity.**

Keywords*— *feature extraction, classification, bag-of-visual-words, SVM, AlexNet, fine-tuning

I. INTRODUCTION

Content-based image retrieval (CBIR) [1] has been one of the most pronounced research areas in computer vision field over the last decade. CBIR is a field which uses visual contents for the retrieval of images from a large digital image database. The rapidly growing number of radiographic equipment in various medical centres and hospitals is making it increasingly difficult for effective retrieval of medical images because of the large number of images captured on daily basis. Therefore, there is a need for an effective and efficient system for information retrieval from the numerous collections stored in the databases [2]. Automatic classification of images can be used to assist in filtering out the irrelevant images and improve the retrieval quality of medical database. Classification is the assignment of images into pre-defined categories or classes [3]. For an effective and accurate classification task, it is of necessity to determine which visual features best represent the image for extraction. Various methods have been employed for the extraction of medical image visual features including Fourier Descriptors, Local binary pattern, moments, edge detection, and Gray Level Co-occurrence Matrix (GLCM) [4-8]. The advances of SIFT and other local features has led researchers to employ the use of Bag-of-Visual-Words (BoVW) model in medical image classification and retrieval task [9-14]. Most recently, with the discovery of deep learning [15-17], machine learning [18-20] witnessed a breakthrough technique that employs deep architecture to model high-level abstractions in data. Convolutional Neural Networks (CNNs) are one of the deep learning architectures which has been used for decades in the field of computer vision [15, 21, 22]. This paper employs both bag-of-visual word and CNN for the task of medical X-ray image automatic classification. It also compares the classification accuracy of using BoVW method for feature extraction with the employment of CNN on the same dataset, ImageCLEF 2007 [23].

II. RELATED WORK

A. Bag-of-Visual-Words Applied on Medical Image Classification and Retrieval

A group of researchers [12] explored various parameters effect on system performance with the use of BoVW. The extraction of the patches was carried out by a regular grid on a large set of images without considering their labels. This was followed by normalization of the patches by subtracting the mean gray level by and then dividing by the standard deviation. Support Vector Machine (SVM) was used to train a classifier on 10677 images while 2000 images were used for testing. The system was applied to categorize chest X-ray pathological-level and was able to differentiate between healthy and diseased cases. Normalized raw data was used for the evaluation and the report showed a high result in the orientation identification and organ identification in ImageCLEF 2009. It also produced a top performance in ImageCLEF 2008 visual based system. Rates of 100.00% sensitivity and 99.92% specificity were recorded in the detection of chest X-ray images from various body parts.

In a research work, Zare [24] attempted to solve the ignorance of spatial information and the ambiguity experienced

978-1-5386-3812-5/18/$31.00 ©2018 IEEE

with the use of BoVW. They employed the use of Probabilistic Latent Semantic Analysis (PLSA) proposed by Hofmann [25] and discriminative SVM classifier for medical X-ray classification. BoVW was extracted and directly fed into SVM classifier for the construction of the classification model in the discriminative approach. In the hybrid generative/discriminative approach, the researchers computed PLSA-based representations of images by fitting into PLSA model, the extracted BoVW. A set of experiments was carried out using ImageCLEF 2007 [26]. The classification result showed there was an improvement compared to the 90% accuracy obtained with the use of bags of words**.** There are **o**ther related works [27-29] that employed BoVW for medical image classification.

B. Convolutional Neural Networks Applied on Medical Image Classification and Retrieval

CNN [30] is one of the deep learning networks known as the most suitable for medical image processing. Bar, Y., et al. [31], explored the strength of CNN to identify chest X-ray images pathologies. In their research work, a non-medical learning was employed as their deep learning approach because of the inadequate labelled training data peculiar to the medical domain. A CNN trained with ImageNet [30, 32] was used and the algorithm was tested on 93 frontal chest X-ray images. The extracted features which serves as the main descriptor were gotten from Decaf implementation of a CNN [30, 33] while the second baseline descriptor used for the research is Picture Codes (PiCoDes) [32] obtained from optimization of ImageNet dataset subset of about 70000 images. These two baseline descriptors were combined (fused together) to produce the best performance. The research benchmark is the testing of some other common descriptors which includes the local Binary pattern (LBP) and the Gist descriptor [34]. The extracted features values were standardized except for the binary values. The work examines three accuracy measurements which were specificity, sensitivity and the area under the ROC curve (AUC). The classification result shows that combining two deep learning extracted features (Decaf) and PiCoDes produced the best performance for all the examined conditions.

The use of three types of CNNs of various capabilities was examined to explore the essential of network depth in modality classification of medical images in a research work [35]. A six weight layers CNN was initially trained from scratch to capture domain specific information using the Glorot uniform [36] as weight initialization. This was followed by the capturing of generic and domain-specific features with the aid of a transfer learning framework. Most layers of VGGNet-16 [37] and ResNet-50 [38] initially pre-trained on ImageNet were fixed on the network while the last fully connected layer was replaced with 30 neurons for the output of 30 posterior probabilities and retrained. The original dataset used was that of ImageCLEF 2015 and 2016 subfigure classification datasets, an additional two sets of augmented data was gotten from ImageCLEF 2013 modality classification task and real-time data augmentation that includes four different transformations. The proposed model was developed through a voting system [39] which receives the computed intensities for each modality and produced the final intensity by combining the outputs of the three CNNs. The system was evaluated on ImageCLEF 2015 and 2016 subfigure classification task producing an improved performance of 76.78% and 86.92% respectively compared to the baselines which produced 60.91% and 85.38% respectively on the same datasets.

The fusion of domain transferred CNNs with a local image dictionary was proposed in [40]. The research work used dictionary-derived sparse spatial pyramid (SSP) [41] to solve the limitation of employing only domain transferred-CNN (DT-CNN) for X-ray image classification. They observed that DT-CNN alone, does not capture the inherent spatial characteristics and local features of X-ray images [42]. The proposed method was a late fusion based method that used a pre-trained CNN of the VGG-19 architecture as a feature extractor for medical X-ray image classification. A vector of 4096 dimensions, which was the output of the last fully-connected layer of VGG-19 was combined together with SSP extracted from image local patches which served as a feature vector to train a linear SVM classifier. The SSP was derived by the initial quantification into visual words, SIFT descriptor combined with the BoVW extracted from image local patches. This was followed by a spatial pooling [43] to learn the local descriptors spatial order. With the use of sparse coding for vector quantization instead of the popular techniques like K-means algorithm, more salient information was captured for image representation. The late fusion approached employed in the research work was able to exploit specific X-ray images local characteristics and the information gotten from DT-CNN.

III. METHODOLOGY

A. BoVW + SVM

For the BoVW feature extraction, SIFT [44] was used as the image descriptor. SIFT uses a keypoint detector in the location-scale space based on the identification of interest points. K-means clustering method was employed for the code book construction and the visual vocabulary term used as cluster centre. This helps to identify the image content collection set that reflects the image visual patterns. After the identification of cluster centres, the frequency of the words appearing in an image is counted and each image is represented as cluster centres. The nearest neighbour with a Euclidean metric was used to assign each feature vector in an image to a cluster centre. At the model generation stage, the extracted BoVW was used as input to support vector machine (SVM) to generate a classification model was evaluated at the training stage. The next step was the application of the model generated on the BoVW representation extracted from a test data for the classification task. Fig. 1 shows the architecture of BoVW + SVM model.

B. CNN + SVM

Transfer learning between two task domains is desirable in medical image classification due to the fact that medical image datasets are smaller compared to the dataset on natural images. For CNN to reach their full potential, a large-scale dataset is required. ImageCLEF 2007 used in this work provides only thousands of label medical images compared to more than 1200000 images present in the ImageNet dataset. Deep CNN takes several weeks to be trained on a multiple GPU on ImageNet. This training duration has led to people releasing their network to benefit others who may want to use if for feature extraction or fine-tuning. AlexNet [30] was used in this work as a feature generator. The output of the fully-connected layer fc6 of AlexNet which was a vector of 4096 dimensions was used as a feature vector to train a linear SVM classifier for the classification task. Fig. 2 shows the architecture of the CNN + SVM model.

C. Fine-tuning of AlexNet

Fine-tuning is a process of training a CNN using a pre-trained weights [45]. AlexNet pre-trained on ImageNet is specifically used for the fine-tuning task. The last fully-connected layer of AlexNet was replaced with a new fully connected layer with 116 neurons. The number of neurons (116) corresponds to the number of classes we require in ImageCLEF 2007 dataset. All the other layers of AlexNet were retained to reserve the generic information which are low level image features [45], while we retrained from scratch the last fully-connected layer to capture domain-specific features. The weights of the last fully-connected layer were initialized but all other layers of the networks were fixed. Fig. 3 shows the architecture of the fine-tuning process.

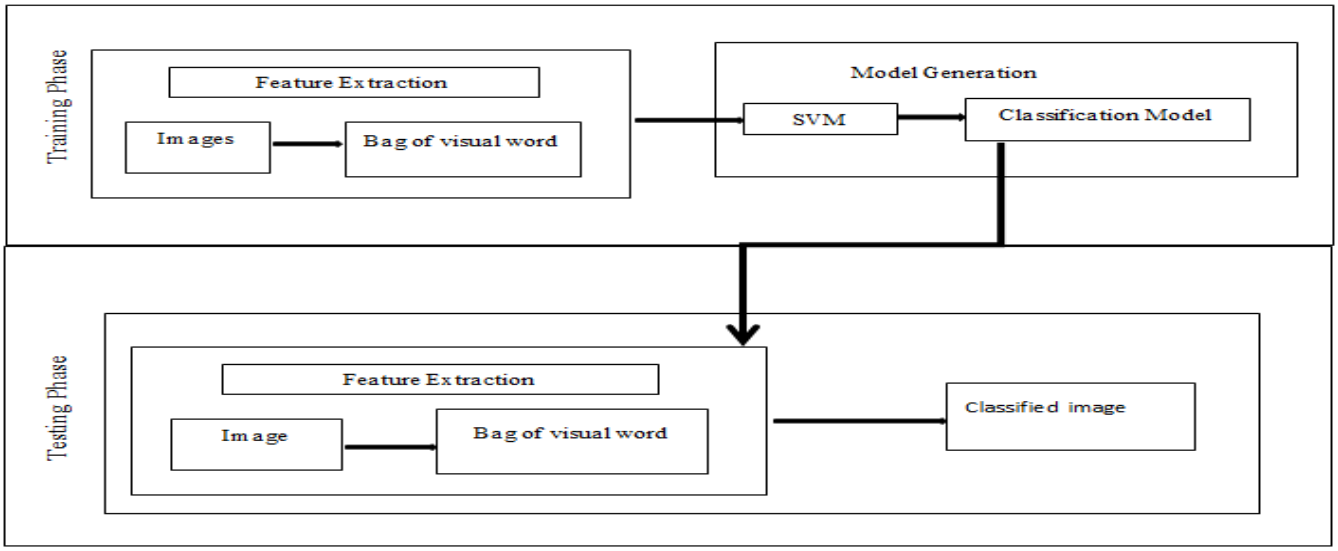

Fig. 1. Classification using BoVW + SVM

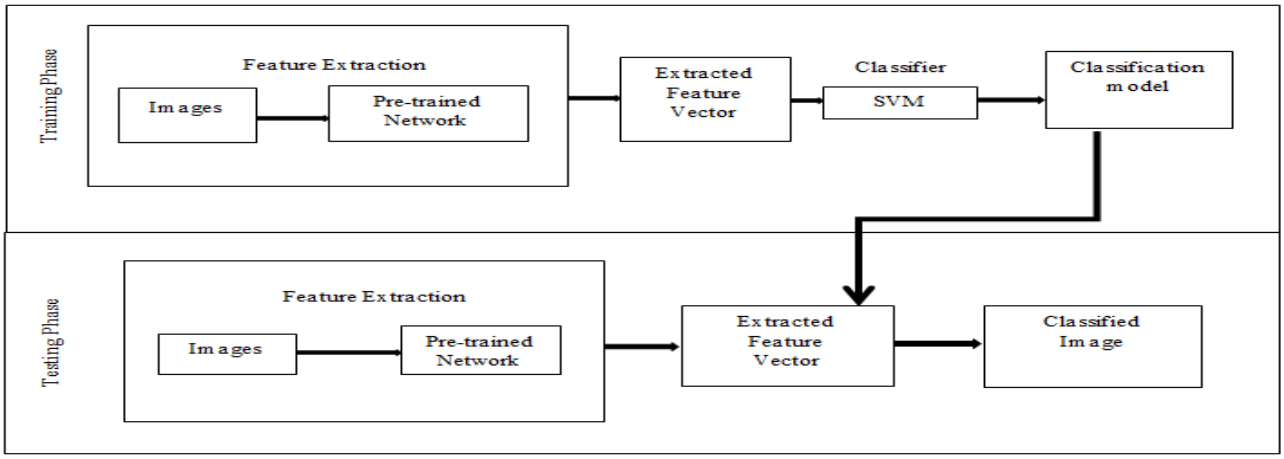

Fig. 2. Classification using Pre-trained Neural Network + SVM

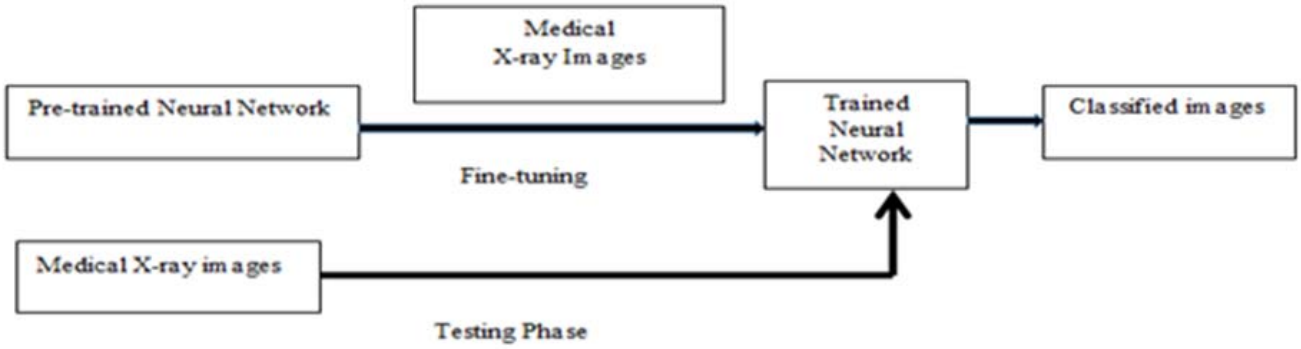

Fig.3. Classification using a fine-tuned network

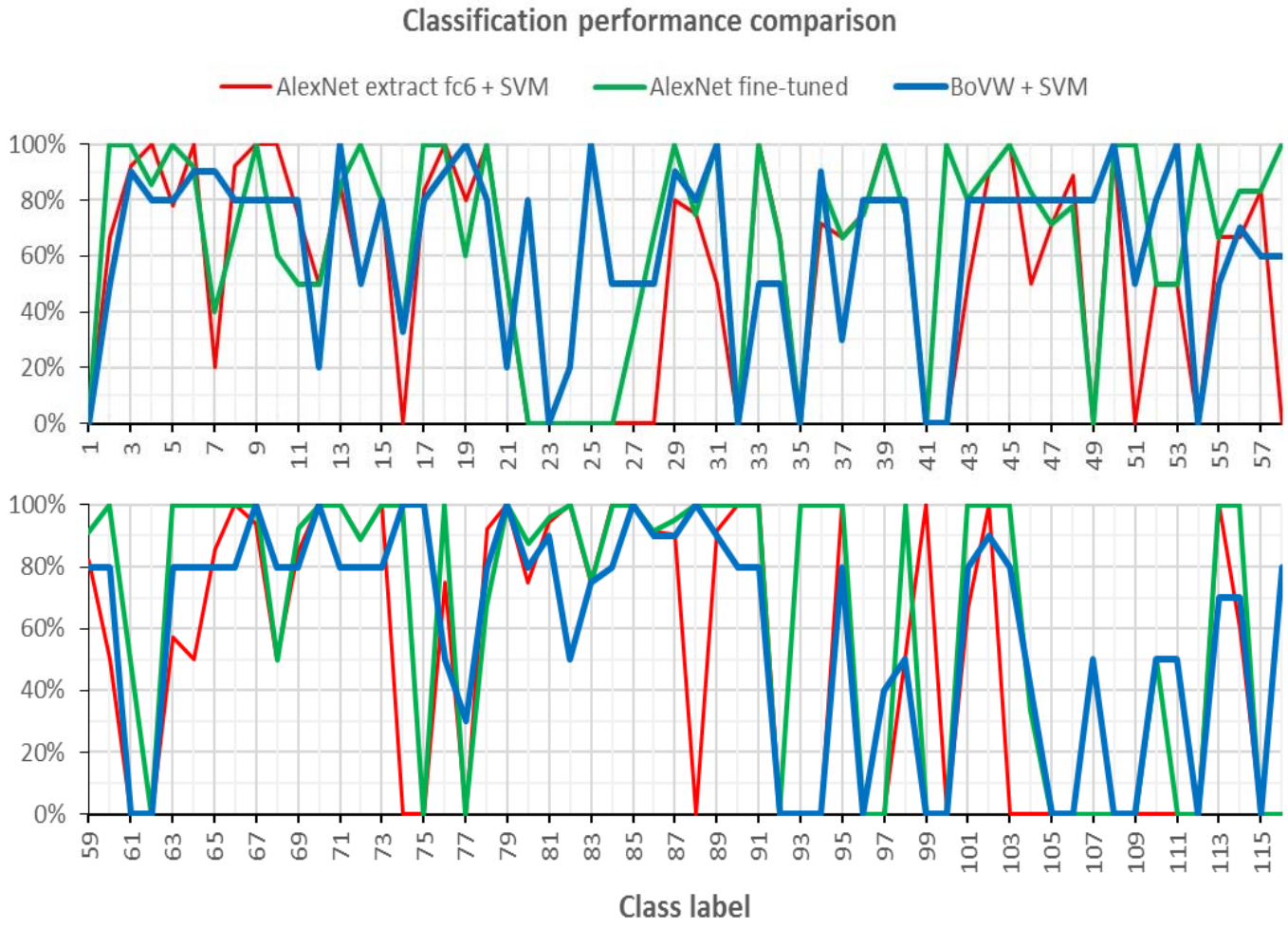

Fig. 4 Comparison of the classification results of the three techniques.

IV. Experimental Results and Discussion

In this section, we carried out an experiment to evaluate the classification performance obtained using the three classification techniques employed in this work. The database used for this research was prepared for the ImageCLEF 2007 automatic annotation task for medical images. The database is made up of 11000 X-ray images labelled with 116 categories of images which are of different modality, body orientation, examined region, and biological system. 1000 unlabelled images were provided for testing. Fig.4 shows the classification result of all the techniques on the same dataset. We observed that from the 116 image classes, the accuracy rate of 24 classes were above 80% when using BoVW + SVM while that of AlexNet + SVM and Fine-tuned AlexNet were 26 and 60 classes respectively. Thus, fine-tuned AlexNet outperformed the two other techniques. We also took note of the image classes whose performance accuracy were below 60%. The BoVW technique produced 43 classes while AlexNet + SVM and Fine-tuned AlexNet produced 33 and 26 classes respectively. For the misclassified images, it was observed that the misclassified classes are within the same sub-body region (Arm category). All the three techniques employed for classification were unable to classify them accurately. Further investigations into the misclassified classes showed that majority of them are suffering from high intra-class variabilities and inter-class similarity. We also discovered that there was no uniform distribution of the number of training images in the sub-body region with high inter-class similarity.

V. Conclusion

As we have seen from the classification results obtained from employing three different classification techniques, it is difficult to obtain a high accuracy for individual class. This is a result of the problem of high inter-class similarity and intra-class variability that exist especially within the image classes from the same sub-body region. This is a common problem when dealing with a large medical database. Fine-tuned AlexNet was able to produce an overall classification accuracy of 86.47% across all 116 classes which is the highest. All the three techniques are competitive based on the experimental results on the entire dataset. They have a difference of about 2% classification accuracy between them. It should be noted that every individual class may not achieve this result. In our future work we may consider CNN architecture such as VGG as a feature generator and also observe the classification results if VGG pre-trained network is fine-tuned on the same dataset. Using a hierarchical classifier can also be considered in attempt to rectify the problems encountered with the misclassified images.

References

[1] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 22, no. 12, pp. 1349-1380, 2000.

[2] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data," *Journal of digital imaging,* vol. 26, no. 6, pp. 1025-1039, 2013.

[3] M. R. Zare, C. Woo, and A. Mueen, "Automatic Classification of medical X-ray Images," *Malaysian Journal of Computer Science,* vol. 26, no. 1, pp. 9-22, 2013.

[4] A. N. J. Raj and V. G. Mahesh, "Zernike-Moments-Based Shape Descriptors for Pattern Recognition and Classification Applications," in *Advanced Image Processing Techniques and Applications*: IGI Global, 2017, pp. 90-120.

[5] L. Houam, A. Hafiane, A. Boukrouche, E. Lespessailles, and R. Jennane, "Texture characterization using local binary pattern and wavelets. Application to bone radiographs," in *Image Processing Theory, Tools and Applications (IPTA), 2012 3rd International Conference on*, 2012, pp. 371-376: IEEE.

[6] H. J. Nussbaumer, *Fast Fourier transform and convolution algorithms*. Springer Science & Business Media, 2012.

[7] A. Rajaei, E. Dallalzadeh, and L. Rangarajan, "Symbolic representation and classification of medical X-ray images," *Signal, Image and Video Processing,* vol. 9, no. 3, pp. 715-725, 2015.

[8] J. Zhang, G.-l. Li, and S.-w. He, "Texture-based image retrieval by edge detection matching glcm," in *High Performance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on*, 2008, pp. 782-786: IEEE.

[9] W. Yang, Z. Lu, M. Yu, M. Huang, Q. Feng, and W. Chen, "Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single-and multiphase contrast-enhanced CT images," *Journal of digital imaging,* vol. 25, no. 6, pp. 708-719, 2012.

[10] M. Srinivas, R. R. Naidu, C. S. Sastry, and C. K. Mohan, "Content based medical image retrieval using dictionary learning," *Neurocomputing,* vol. 168, p. 880, 2015.

[11] M. Reza Zare, W. Chaw Seng, and A. Mueen, "Automatic classification of medical X-ray images using a bag of visual words," *IET Computer Vision,* vol. 7, no. 2, pp. 105-114, 2013.

[12] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Transactions on Medical Imaging,* vol. 30, no. 3, pp. 733-746, 2011.

[13] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, "Histopathology image classification using bag of features and kernel functions," in *Conference on Artificial Intelligence in Medicine in Europe*, 2009, pp. 126-135: Springer.

[14] R. Xu, Y. Hirano, R. Tachibana, and S. Kido, "Classification of diffuse lung disease patterns on high-resolution computed tomography by a bag of words approach," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011,* pp. 183-190, 2011.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, no. 7553, pp. 436-444, 2015.

[16] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks,* vol. 61, pp. 85-117, 2015.

[17] L. Deng and D. Yu, "Deep learning: methods and applications," *Foundations and Trends® in Signal Processing,* vol. 7, no. 3–4, pp. 197-387, 2014.

[18] X. Meng *et al.*, "Mllib: Machine learning in apache spark," *The Journal of Machine Learning Research,* vol. 17, no. 1, pp. 1235-1241, 2016.

[19] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Machine learning,* vol. 3, no. 2, pp. 95-99, 1988.

[20] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, "An introduction to MCMC for machine learning," *Machine learning,* vol. 50, no. 1-2, pp. 5-43, 2003.

[21] K. Fukushima and S. Miyake, "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition," in *Competition and cooperation in neural nets*: Springer, 1982, pp. 267-285.

[22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.

[23] H. Müller, T. Deselaers, T. M. Deserno, J. Kalpathy-Cramer, E. Kim, and W. R. Hersh, "Overview of the ImageCLEFmed 2007 Medical Retrieval and Medical Annotation Tasks," in *CLEF*, 2007, pp. 472-491: Springer.

[24] M. Reza Zare, W. Chaw Seng, A. Mueen, and M. Awedh, "Automatic classification of medical X-ray images: hybrid generative-discriminative approach," *IET Image Processing,* vol. 7, no. 5, pp. 523-532, 2013.

[25] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine learning,* vol. 42, no. 1, pp. 177-196, 2001.
[26] H. Müller *et al.*, "Overview of the CLEF 2009 medical image retrieval track," in *Workshop of the Cross-Language Evaluation Forum for European Languages*, 2009, pp. 72-84: Springer.
[27] L. Valavanis, S. Stathopoulos, and T. Kalamboukis, "Fusion of Bag-of-Words Models for Image Classification in the Medical Domain," in *European Conference on Information Retrieval*, 2017, pp. 134-145: Springer.
[28] V. González-Castro, M. d. C. V. Hernández, P. A. Armitage, and J. M. Wardlaw, "Automatic rating of perivascular spaces in brain MRI using bag of visual words," in *International Conference Image Analysis and Recognition*, 2016, pp. 642-649: Springer.
[29] M. R. Zare, A. Mueen, and W. C. Seng, "Automatic classification of medical X-ray images using a bag of visual words," *IET Computer Vision,* vol. 7, no. 2, pp. 105-114, 2013.
[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.
[31] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification," in *Proc. SPIE*, 2015, vol. 9414, p. 94140V.
[32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248-255: IEEE.
[33] J. Donahue *et al.*, "Decaf: A deep convolutional activation feature for generic visual recognition," in *International conference on machine learning*, 2014, pp. 647-655.
[34] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision,* vol. 42, no. 3, pp. 145-175, 2001.
[35] Y. Yu, H. Lin, J. Meng, X. Wei, H. Guo, and Z. Zhao, "Deep Transfer Learning for Modality Classification of Medical Images," *Information,* vol. 8, no. 3, p. 91, 2017.
[36] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249-256.
[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.
[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015.
[39] L. I. Kuncheva and J. J. Rodríguez, "A weighted voting framework for classifiers ensembles," *Knowledge and Information Systems,* vol. 38, no. 2, pp. 259-275, 2014.
[40] E. Ahn, A. Kumar, J. Kim, C. Li, D. Feng, and M. Fulham, "X-ray image classification using domain transferred convolutional neural networks and local sparse spatial pyramid," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, 2016, pp. 855-858: IEEE.
[41] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 37, no. 9, pp. 1904-1916, 2015.
[42] S. Choi, "X-ray Image Body Part Clustering using Deep Convolutional Neural Network: SNUMedinfo at ImageCLEF 2015 Medical Clustering Task," in *CLEF (Working Notes)*, 2015.
[43] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Computer vision and pattern recognition, 2006 IEEE computer society conference on*, 2006, vol. 2, pp. 2169-2178: IEEE.
[44] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision,* vol. 60, no. 2, pp. 91-110, 2004.
[45] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE transactions on medical imaging,* vol. 35, no. 5, pp. 1299-1312, 2016.