Machine Learning Based Common Radiologist-Level Pneumonia Detection on Chest X-rays

Min-Jen Tsai and Yu-Han Tao
National Chiao Tung University, Institute of Information Management, 1001 Ta-Hsueh Road, Hsin-Chu, 300, Taiwan, R.O.C.

Abstract
For digital pathology, automatic recognition of different tissue types in histological images is important for diagnostic assistance and healthcare. Since histological images generally contain more than one tissue type, multi-class texture analysis plays a critical role to solve this problem. The chest X-ray is one of the most commonly accessible radiological examinations for screening and diagnosis of many lung diseases. This study examines the important statistical features including use the Convolutional Layer, ReLU Layer and Pooling Layer for detect pneumonia from Chest X-rays identification by using Convolutional Neural Network (CNN) and decision fusion of feature selection. Our algorithm, is a 144 layers Convolutional Neural Network (CNN) trained on Chest X-rays 14 diseases, currently the largest publicly available Chest X-rays dataset, containing over 100,000 frontal view Chest X-rays classification of histological images with 14 diseases. We detect all 14 diseases in Chest X-rays and achieve state of the art results on all 14 diseases. The average experimental results achieves high identification rate which is significantly superior to the existing known methods. In summary, the proposed method based on machine learning outperforms the techniques described in the literatures and achieve high classification accuracy rate at 80.90% for 144 layers of Convolutional Neural Network (CNN) which demonstrates promising applications for Chest X-rays classification of histological images.
Keywords *Classification, Support Vector Machine, Convolutional Neural Network (CNN), Machine Learning, Deep Learning.*

Introduction

More than 1 million adults are hospitalized with lung diseases and around 50,000 die from the disease every year in the US alone (CDC, 2017), pneumonia is one type of lung diseases that is frequently evaluated clinically. Chest X-rays are currently the best available method for diagnosing lung diseases (WHO, 2001), playing a crucial role in clinical care (Franquet, 2001) and epidemiological studies (Cherian et al., 2005). However, detecting pneumonia in Chest X-rays is a challenging task that relies on the availability of expert radiologists. It is imperative to identify multiclass lung diseases based on tissue types of Chest X-rays histological images. In this work, we present a model that can automatically detect pneumonia from Chest X-rays at a level exceeding practicing radiologists.

The rapid and tremendous progress has been evidenced in a range of computer vision problems via deep learning and large-scale diseases histological images datasets [1, 2, 3, 4]. Drastically improved quantitative performances in object recognition, detection and segmentation are demonstrated in comparison to previous shallow methodologies built upon hand-crafted image features. Deep neural network representations further make the joint language and vision learning tasks more feasible to solve, in image captioning [5, 6, 7, 8, 9], visual question answering [10, 11, 12, 13] and knowledge-guided transfer learning [14, 15], and so on.

Among previous works, several methods for histological texture diseases analysis have been studied. Kather et.al [16] identified histological images of human colorectal cancer including eight different types of tissue by using texture descriptors and several classifiers. They divided ten original tissue images into patches for identification based on classes. Mattfeldt, et al. [17] studied the correlation between epithelial cells and lumina from low grade to high grade prostatic cancer progression in terms of the Gleason score. They implemented multiclass pattern recognition constructed by spatial statistics, as contrasting to the mathematical method of binary pattern recognition. Huang and Lee [18] examined variations of intensity and texture complexity for histological grading of prostate tissues using two feature extraction methods based on fractal dimension. Several feature filter sets were applied and also used by the sequential floating forward selection method to optimize the classification results.

Furthermore, Signolle et.al [19] segmented histopathology slides to identify various types of ovarian carcinoma stroma using wavelet-domain hidden Markov tree model and a pairwise classifiers design and selection. Additionally, Yang et al. [20] applied a grid-enabled decision support system to perform automatic analysis of breast tissue microarray images. Four different types of filter banks were applied to classify several major subtypes of breast cancer using k-nearest neighbor (kNN) and decision tree integrated into a Bayesian framework. Additionally, Xiaosong et.al [21] using natural language processing to perform automatic analysis of frontal view X-ray images of 32,717 unique patients with the text mined eight disease image labels (where each image can have multi-labels), demonstrate that these commonly occurring thoracic diseases can be detected and even spatially-located via a unified weakly supervised multi-label image classification and disease localization framework.

Previous studies commonly extracted the statistical features [21, 22] from the printed documents by Local Binary Pattern (LBP), Gray-Level Co-Occurrence Matrix (GLCM), Discrete Wavelet Transform (DWT), spatial filters…etc. for feature extraction. Subsequently, support vector machine (SVM) [23] is adopted for the forensic classification system. However,

978-1-7281-2194-9/19/$31.00 ©2019 IEEE

above mentioned approaches require significant human involvement with expert knowledge during the procedures of feature extraction, feature selection and classification.

This paper intends to leverage the technology advancement of deep learning for Chest X-rays tissue identification, which can reduce the burden of human machine interaction with automatic classification capability. To achieve those goals, the objective of this paper is to obtain the best performance for Chest X-rays tissue identification for histopathology images where the deep learning technique is utilized. Furthermore, the aim of the research is also to compare the best decision results from machine learning with the existing techniques.

Machine learning is a subfield of artificial intelligence (AI) [24]. Convolutional neural networks (CNNs) [25] recently have shown to be very effective in complex image classification tasks in machine learning field. The main benefit of using CNNs with respect to traditional fully-connected neural networks is the reduced amount of parameters to be learned. Convolutional layers made of small size kernels are an effective way to produce high-level features that are fed to fully-connected layers and significantly reduce the computation time with high accuracy rate.

To achieve those goals, the objective of this paper is to obtain the best performance for Cancer Tissue Classification where the deep learning technique is utilized. Our model, is a 144 layers Convolutional Neural Network (CNN) trained on Chest X-rays 14 diseases, currently the largest publicly available Chest X-rays dataset, containing over 100,000 frontal view Chest X-rays classification of histological images with 14 different thoracic diseases, including pneumonia. Detecting pneumonia in chest radiography can be difficult for radiologists. The appearance of pneumonia in Chest X-rays images is often vague, and can mimic many other benign abnormalities. We detect all 14 diseases in Chest X-rays and achieve state of the art results on all 14 diseases. Furthermore, the aim of the research is also to compare the best decision results from machine learning with the existing techniques.

As a consequence, this paper is organized as following: Section 2 describes the related works, classification techniques, deep learning methodology and feature based approach. Section 3 presents the proposed technique used in this study. In Section 4, experimental results are reported with discussion, and Section 5 concludes the paper with areas of possible future investigation.

Related Works

The structures of human lung diseases comprise several tissue types that are able to be distinguished by histopathological evaluation of Chest X-rays histological images stained tissue sections. Table 1 summarized research papers on the topic of authenticating the structures of human histopathological images.

This study examines the important statistical features including use the Convolutional Layer, ReLU Layer and Pooling Layer for Chest X-rays identification by using Convolutional Neural Network (CNN) and decision fusion of feature selection. The average experimental results achieves high identification rate which is significantly superior to the existing known methods. In summary, the proposed method based on machine learning outperforms the techniques described in the literatures and achieve high classification accuracy rate at 80.90% for 144 layers of neural network which demonstrates promising applications for Chest X-rays identification classification of histological images.

Table 1 Research papers on the topic of authenticating the structures of human histopathological images

Research	Filters	Research object	Classifier	Claimed accuracy rate
[16]	LBP, Gabor, GLCM	CRC for eight classes	SVM	87.4 %
[19]	Wavelet	Tumor epithelium	Hidden Markov tree	71.5 %
[20]	Four different filter sets	Breast cancer	kNN, Bayesian, C4.5 decision tree, and SVM	89 %
[26]		Breast Cancer	CNN	90 %
This study		Chest X-rays	CNN	80.90 %

Classification Techniques

SVM [23] can classify pixels for images according to textural cues. The best hyperplane separation among them can be found by measuring the margin hyperplane and looking for maximum points [27]. In addition, SVM is able to obtain the best result in comparison among feature extractions for a multi texture classification problem [28].

Other useful classifiers worth mentioning are Gaussian mixture model (GMM) and AdaBoost algorithm. Gaussian mixture model (GMM) [26] is a probabilistic model for representing normally distributed subpopulations within an overall population. GMMs had been very successful in modeling speech features and in acoustic modeling for speech recognition for many years (until around year 2010–2011 when deep neural networks were shown to outperform the GMMs) [26].

The AdaBoost algorithm, which is one of the most popular boosting algorithms, generates a sequence of base models with different weight distributions over the training set [29]. AdaBoost is sensitive to noisy data and outliers. In some problems, it can be less susceptible to the overfitting problem than other learning algorithms. However, in [30, 31], AdaBoost algorithm are not always expected to improve the performance of SVMs, and even they worsen the performance particularly.

From Table 1, the classifiers of the reported highest accurate rates for image documents are either from feature based SVM or CNN model of machine learning for Research papers on the topic of authenticating the structures of human tumors. Unlike feature based situation, deep learning technique allows the proposed model to learn the distribution and train the system automatically. Few human intervention is involved for the whole architecture for accurate identification rate which is technically inspiring for researchers to enhance the existing techniques and explore new forensics approaches.

Machine Learning and Deep Learning

Machine learning is mainly about how the computer simulates and realizes human learning behavior, and allowing machine to do self-learning and acquire new knowledge or skills. The general procedures applying machine learning to solve the above mentioned problem is showed in Fig. 2. Reasoning, prediction or identification are the target for the machine learning. Data preprocessing, feature extraction and feature selection play key roles for machine learning.

Accordingly, it is important to understand how the human brain works, and Hubel et al. [30] had found that the operation of the neuron system is hierarchical based on the functional analysis of the cortex cells of the cat to find the corresponding relationship between neurons. More advanced machine learning system called the deep learning has used more hidden layers of artificial neural network as shown in Fig. 3, to achieve the multi-layer operation of the neural system. The basic concept is the output of the previous layer as the input of the next layer and the input information can be hierarchically expressed by combining the low-level features to form more high-level abstract features.

Hinton et al. [31] have reported that the multi-hidden artificial neural network has excellent learning capability. Consequently, deep learning can establish a multi-hidden layer of learning model to process a large number of training samples, learn useful features and enhance the classification or prediction accuracy. Deep learning has also developed many different models such as: Auto Encoder, Sparse Coding, Restricted Boltzmann Machine (RBM) [32], Deep Belief Networks and Convolutional Neural Networks (CNN). Related studies have achieved great success in many fields of human machine interaction [33].

Fig. 2 The learning procedures of machine learning

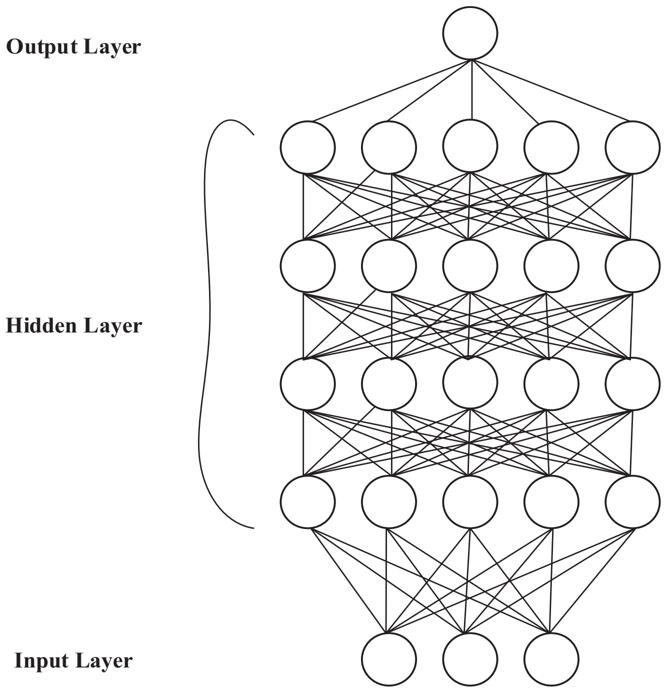

Fig. 3 The multi hidden layer structure of deep learning model

Convolutional neural networks (CNN)

Convolutional neural networks (CNN) is a type of artificial neural network that was proposed by Lecun et al. [33] it has become one of the popular tools in the field of speech analysis and image recognition. Its shared weights network structure is very similar to the actual biological neural network in simulation. This feature can also reduce the complexity of network model and reduce the number of parameters. The neurons in the convoluted neural network are derived from the concept of receptive field proposed by Hubel et al. [30] through the study of cat visual cortical cells, followed by Japanese scholar Kunihiko Fukushima. Based on the concept of receptive field, the neocognition can be regarded as the first realization network of the convolutional neural network and the first application of the field in the artificial neural network. The neural cognition machine usually contains two types of neurons: the S-element of the characteristic extraction and the anti-deformation C-element. There are two important parameters in the S-element: the field and the threshold parameter, the former determines the number of input connections, the latter controls the degree of response to the characteristic sub-mode. Thereafter, the neural cognitive machine has more different development, convoluted neural network can be seen as a form of neural cognitive promotion, and neural cognitive machine is a special case of convolution neural network.

Convolution neural network [1] is a multi-layer neural network that each layer consists of multiple two-dimensional planes. This plane is composed of multiple independent neurons, the basic concept of network architecture as shown in Fig. 4. The C layer in the graph is the feature extraction layer, and the input of each neuron connects with the local receptive field of the previous layer and the local feature is extracted. The S layer in the graph is the feature mapping layer, and each feature layer of the network is composed of multiple feature mapping layers. Each feature is mapped to a plane, and the weights of all the neurons on the plane are equal. In addition, since neurons on a map surface share weights, it is possible to reduce the number of network parameters and reduce the complexity of network parameter selection. Each feature extraction layer (C layer) in the convolutional neural network is followed by a computational layer (S layer) for local averaging and secondary extraction.

The training of convolutional neural network is mainly operated through the back propagation algorithm and stochastic gradient descent with momentum algorithm. Back propagation can be divided into two phases: propagation and weight update. It technically calculates the gradient of the loss function, and then feedback to the optimization method for weight update[34]. Compared to the standard gradient descent method using the entire dataset, the stochastic gradient descent method for each period of interaction uses the training subset called mini-batch. It minimizes the error function by updating parameters of weight and bias, and the parameter updating is calculated by the gradient of the loss function fed back by the Back Propagation [34].

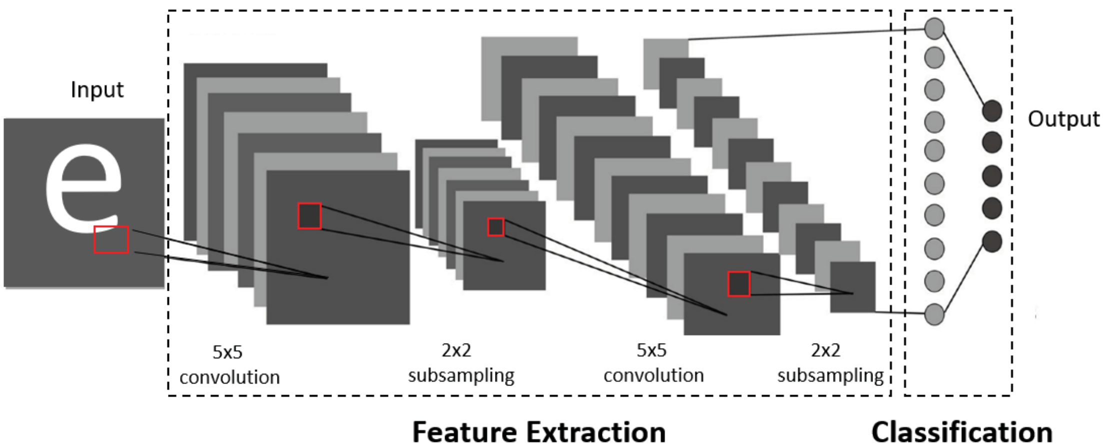

Fig. 4 The network architecture of CNNs

The Deep Learning Based Classification Approach

A. The spatial features

Feature extraction and feature selection and classification are all substituted by the convolutional neural network. The different combination of convolutional layer, ReLU layer and pooling layer form a multi-layer neural network. Apply the system to train samples, automatically learn the characteristics of the images from Chest X-rays histological images. The fully-connected layer with soft-max operation play the role of classifier.

B. **Network model selection**

This study considers Convolutional Neural Network (CNN) as experts and feature sets as alternatives. The text and image files of unknown classification to the trained neural network model in combination with Fully-Connected Layer and Soft-max Layer. The classifier will analyze the classification of the Chest X-rays histological images based on the mark of the training result. According to the identification result, the Chest X-rays histological images can be classified.

C. **Network layer**

Table 2 tabulates the Convolutional Neural Network (CNN) architecture design, in which $\boldsymbol{CONV + POOL_{max}}$ represents the convolutional layer followed by the use of the maximum generalized pooling layer, and $\boldsymbol{CONV + POOL_{avg}}$ is a pooling layer that follows the generalization of the average.

The systematic framework for Chest X-rays histological images classification based on deep learning has been developed and the flowchart is shown in Fig. 5 with the following procedures.

In this study, we use the selected datasets such as is shown in Fig. 6. It was conducted by Kather, et al. [21]. Implemented by the same experimental setting, 80% Chest X-rays histological images from a tissue type are randomly selected to train the CNN network architecture, whereas at least another 20% images, randomly taken from the same tissue data sets, are tested during the identification step. In the experiments, we segment 14 diseases images application set that are in different tissue types by sub-dividing each image of 10,000 images overlapped at 1024x1024 pixel sizes.

Since feature based classification approach is commonly used, the technique will be compared in this study. The procedures and those image features will be briefly explained.

Table 2 Summary of parameters for CNNs

	Layers				
	1	2	3	4	5
Type	$CONV + POOL_{max}$	$CONV + POOL_{max}$	$CONV + POOL_{avg}$	$CONV + POOL_{avg}$	$CONV + POOL_{avg}$
Weights	7*7*3*64	1*1*64*64	1*1*192*64	1*1*480*192	1*1*832*256
Bias	1*1*64	1*1*64	1*1*64	1*1*192	1*1*256
Convolution Stride	1*1	1*1	1*1	1*1	1*1
Filter Numbers	64	64	64	192	256
Pooling Size	3*3	3*3	3*3	3*3	3*3
Pooling Stride	3*3	3*3	3*3	3*3	3*3
Padding Size	3*3	3*3	3*3	3*3	3*3

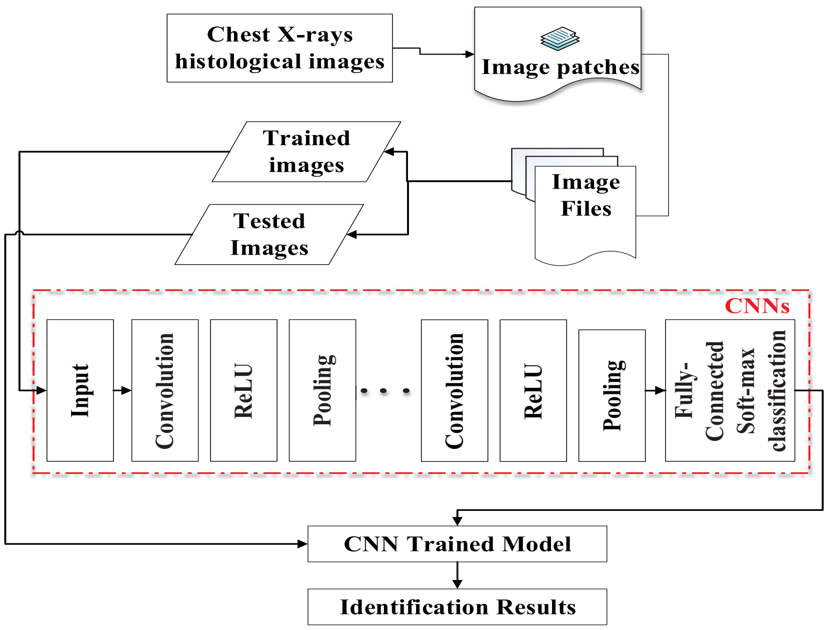

Fig. 5 The flowchart of CNNs system

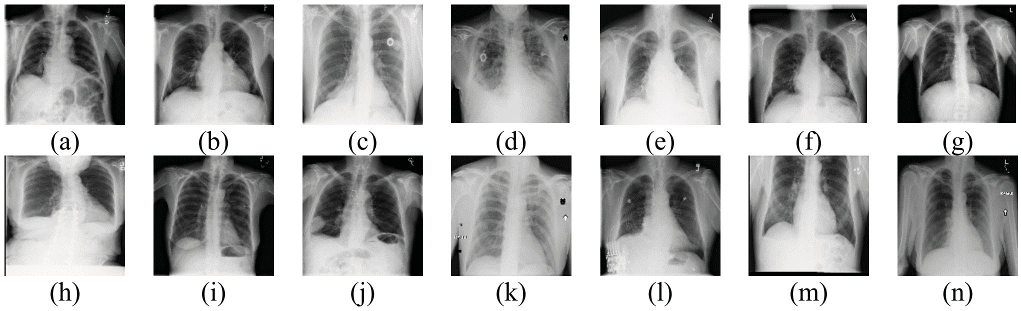

Fig. 6 Image database samples [21] (a) Atelectasis, (b) Cardiomegaly, (c) Effusion, (d) Infiltration, (e) Mass, (f) Nodule, (g) Pneumonia, (h) Pneumothorax, (i) Consolidation, (j) Edema, (k) Emphysema, (l) Fibrosis, (m) Pleural Thickening, (n) Hernia.

Experimental Results

While establishing a deep neural network, the structure will be designed according to the amount of data and the complexity of the image content. The design principle is not only to get a concise network architecture for less training time but also to achieve high accuracy rate.

In this study, the network architecture has five different depth, either using 1, 2, 3, 4 or 5 convolutional layers respectively, and each convolutional layer will be equipped with ReLU layer and pooling layer. Therefore, total 144 layers neural network models are design for comparison. In this stage, the network architecture has three different depth, either using 1, 2, 3, 4 or 5 convolutional layers respectively, and each convolutional layer will be equipped with ReLU layer and pooling layer. Therefore, total 144 layers neural network models are design for comparison. Each Chest X-rays histological images are sent to the 144 layers CNNs network (details in Table I) for training, classified respectively. Since the highest accuracy rate can be achieved by using 144 layers CNNs network from above analysis. After the most important layers have been decided, the tissue type identification can be finally investigated.

In summary, from above analyses, the superior accuracy rates justify the effectiveness of our proposed method for using 144 layers of neural network in identifying Chest X-rays histological images. It demonstrates that our proposed method is superior to the previous studies and the technique can effectively identify the Chest X-rays histological images based histological images.

Discussion

1) Automatic recognition is an essential part in the digital pathology to analyze different tissue types. To achieve the successful studies, multi-disciplinary experts are needed to collaborate among researches such as medical experts, pathologists, computer vision experts, …etc.

2) How to identify the region where different tissue types locate is still a critical issue in digital histopathology in practice. In additional, limited data sources are available online and this study used the dataset from [21], but more medical data are needed to validate its capability in general use.

Conclusion

This paper presents different tissue types in pneumonia from Chest X-rays images classification for pneumonia detection using Convolutional Neural Network (CNN) based decision fusion. It demonstrates that our identification results achieve the best accuracy rates and our proposed method is superior to the previous studies for Chest X-rays images classification. Therefore, the developed technique in this study can effectively identify the Chest X-rays images. The results also confirm that human solid tumors with the complex structures can be distinguished based on the texture of tissue types. The deep learning based classification is very useful for digital pathology.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, pages 115(3): 211–252, 2015.

[3] M. Everingham, S. M. A. Eslami, L. J. Van Gool, C. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision, pages 111(1): 98–136, 2015.

[4] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollr, and L. Zitnick. Microsoft coco: Common objects in context. ECCV, pages (5): 740–755, 2014.

[5] I. Vendrov, R. Kiros, S. Fidler, and R. Urtasun. Orderembeddings of images and language. In ICLR, 2016.

[6] A. Karpathy and L. Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In CVPR, 2015.

[7] B. Plummer, L. Wang, C. Cervantes, J. Caicedo, J. Hockenmaier, and S. Lazebnik. Flickr30k entities:

Collecting regionto-phrase correspondences for richer image-to-sentence models. In ICCV, 2015.
[8] I. Vendrov, R. Kiros, S. Fidler, and R. Urtasun. Orderembeddings of images and language. In ICLR, 2016.
[9] J. Johnson, A. Karpathy, and L. Fei-Fei. Densecap: Fully convolutional localization networks for dense captioning. In CVPR, 2016.
[10] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, and L. Zitnick. Vqa: Visual question answering. In ICCV, 2015.
[11] M. Tapaswi, Y. Zhu, R. Stiefelhagen, A. Torralba, R. Urtasun, and S. Fidler. Movieqa: Understanding stories in movies through question-answering. In ICCV, 2015.
[12] Q. Wu, P. Wang, C. Shen, A. Dick, and A. van den Hengel. Ask me anything: free-form visual question answering based on knowledge from external sources. In CVPR, 2016.
[13] Y. Zhu, O. Groth, M. Bernstein, and L. Fei-Fei. Visual7w:Grounded question answering in images. In CVPR, 2016.
[14] J. Ba, K. Swersky, S. Fidler, and R. Salakhutdinov. Predicting deep zero-shot convolutional neural networks using textual descriptions. In ICCV, 2015.
[15] R. Qiao, L. Liu, C. Shen, and A. van den Hengel. Less is more: zero-shot learning from online textual documents with noise suppression. In CVPR, 2016.
[16] J.N. Kather, C.A. Wei, F. Bianconi, S. M. Melchers, L.R. Schad, T. Gaiser, A. Marx, and F.G. Zöllner, "Multi-class texture analysis in colorectal cancer histology", Scientific Reports 6, Article number: 27988, 2016.
[17] T. Mattfeldt, P. Grahovac, and S. Luck. Multiclass Pattern Recognition of the Gleason Score of Prostatic Carcinomas Using Methods of Spatial Statistics. Image Anal Stereol 32, pp. 155–165, 2013.
[18] P.W Huang & C.H Lee, Automatic classification for pathological prostate images based on fractal analysis. IEEE T Med Imaging 28, pp. 1037–1050, 2009.
[19] N., Signolle, M. Revenu, B. Plancoulaine, and P. Herlin, Wavelet-based multiscale texture segmentation: Application to stromal compartment characterization on virtual slides. Signal Process 90, pp. 2412–2422, 2010.
[20] L. Yang, Virtual microscopy and grid-enabled decision support for large-scale analysis of imaged pathology specimens. IEEE T Med Imaging 13, pp. 636–644, 2009.
[21] R.C. Gonzales, R.E. Woods, Digital Image Processing, third ed., Prentice Hall, New Jersey, 2008.
[22] R.C. Gonzales, R.E. Woods, S.L. Eddins, Digital Image Processing Using MATLAB, second ed., Gatesmark, United Stated, 2009.
[23] C.W. Hsu, C.C. Chang, C.J. Lin, A Practical Guide to Support Vector Classification, National Taiwan University, Taipei, 2003, http://www.csie.ntu.edu.tw/cjlin/papers/guide/guide.pdf.
[24] Bishop. C.M., Pattern Recognition and Machine Learning, Springer, New York, 2006.
[25] A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet classification with deep convolutional neural networks, Processing of International Conference on Neural Information Processing Systems, NIPS, Vol. 1, 2012, pp. 1097–1105.
[26] D. Yu, L. Deng, Automatic Speech Recognition: A Deep Learning Approach, Springer-Verlag, London, 2015.
[27] S. Tong, D. Koller, Support vector machine active learning with applications to text classification, J. Mach. Learn. Res. (2001) 45–66.
[28] K.I. Kim, K. Jung, S.H. Park, H.J. Kim, Support vector machines for texture classification, IEEE Trans. Pattern Anal. Mach. Intell. 24 (11) (2002) 1542–1550, http://dx.doi.org/10.1109/TPAMI.2002.1046177.
[29] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. System Sci. 55 (1) (1997) 119–139.
[30] D. Hubel, T. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, J. Physiol. 160 (1) (1962) 106–154.
[31] G. Hinton, R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[32] J. Markoff, How Many Computers to Identify a Cat? 16, 000, The New York. Retrieved June 22, 2012, from https://mobile.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html.
[33] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[34] E. Rumelhart, E. Geoffrey, J. Ronald, Learning representations by back-propagating errors, Nature 323 (1986) 533–536.