Advances and Applications in Mathematical Sciences
Volume 18, Issue 8, June 2019, Pages 753-765

SYSTEMATIC REVIEW OF RECENT COMPUTER AIDED DIAGNOSTIC APPROACHES IN MEDICAL IMAGING

SUKHENDRA SINGH and BIPIN KUMAR TRIPATHI

Information Technology Department
JSS Academy of Technical Education
Noida, India
E-mail: sukhendrasingh@jssaten.ac.in

CSE Department
HBTU Kanpur, India

Abstract

Computational Intelligence is impacting many areas of daily life. Health is one of them. One emerging technology of AI, machine learning is being utilized to diagnose the disease present in the patient. This area which deals with computational Intelligence techniques by analyzing scanned images of the affected body part by the suitable imaging device to detect the presence or absence of disease is known as Computer-assisted diagnosis. These scanned images are of very high resolution still it is difficult with just by visual examination for any human to conclude about the disease in the patient. Research is this direction is more than three decades but still not widely popular and acceptable among community of radiologists and physicians. A number of approaches have been suggested for Computer assisted diagnosis for different set of diseases and imaging modalities. CAD systems use trained datasets for classification. These training datasets are prepared by team of experts who manually annotates images. These CAD systems processes the images and highlights image segments where there is greater degree of suspicion of lesions which are further processed by radiologists for final conclusion. Such systems save a lot of time and effort of radiologists and also can increase their efficiency. In this paper, authors have presented systematic literature review of recent approaches by keeping focus on critical performance issues and also presented future scope of machine learning approaches which could be used in computer assisted diagnosis.

I. Introduction

When in many disease, doctors recommend different types of medical examinations of a particular organ of body part from various devices like *X* rays, CT scan, MRI, Ultra sound etc. A detailed description of a particular

2010 Mathematics Subject Classification: 92C55.
Keywords: machine learning, medical imaging, computer assisted diagnosis.
Received February 1, 2019; Accepted March 17, 2019

organ is converted in the form of digital image which is further analyzed to detect absence or presence of disease and related predictions. For this, visual inspection performed by physician is neither enough nor reliable. These images, data set are processed by machine learning techniques to produce results which could be efficiently and reliably interpreted by radiologists and physicians to predict and decide further direction of medication. This discipline which deals with examination of medical images presents new direction of medical imaging. These medical images act as training data which are used by machine learning systems for diagnosis. A field of research that targets this is known as computer assisted diagnosis. Computer assisted diagnosis systems produce every information for radiologists and physicians as well so that they can use it for prediction of disease. But the interpretation of results produced by such CAD system still depends upon, knowledge, skill, experience of radiologist and physicians. These system acts as second opinion for any disease are for their assistance for efficient and reliable diagnosis. Apart from experience and skill of doctors, diagnosis also depends upon suitable selection of modality chosen for test. These imaging modalities can produce much useful information for radiologists. For example CT scan can provide following diagnostic information which may lead some inference [i] Lesions [ii] breaks [iii] hemination [iv] abnormal growth [v] rupture. These modalities can also provide benefits such as [i] determining when surgeries are necessary [ii] minimizing the requirement of exploratory surgeries [iii] improving cancer diagnosis and treatment [iv] reducing the length of hospitalizations. Every type of modality provides different type of medical information that may be useful for diagnosis and cure. Machine learning is used to design systems which can make human like decisions; these systems can also improve their performance with increase in experience like human. These systems can analyze complex patterns and perform association. Before the medical images can be analyzed for prediction, they need to be preprocessed well in order increase performance and reliability. A simple model of computer assisted diagnosis can be visualized as two class classification problem in which tumor in the patent is classified into two groups, malignant or benign. Apart from classifying medical images into categories, CAD systems assist physicians and radiologists in many other ways so that they can concentrate on dominant features.

Remaining paper is divided as follows: part 1 presents background concepts for medical Imaging and machine learning; part 2 presents performance issues and other challenging issues involved in the field of computer assisted diagnosis; part 3 presents review of recent approaches used for disease diagnosis and detection using machine learning; part 4 presents future direction of research which will how medical science is going to change with the revolution in computational intelligence followed by conclusion.

II. Background

Computer assisted diagnosis applies techniques of digital image processing to preprocess and enhance quality of image so that the same can be analyzed efficiently. The analysis part will apply machine learning. The process of capturing images of some body organ through various modalities and then input to systems which predicts diagnosis involves stages consisting of (i) preprocessing of scanned images which includes image registration and enhancement, normalization, (ii) segmentation which divides images into similar regions and identifies regions where there is suspicion of lesions .Theses regions are called region of interest (ROI) also called candidate for lesions (iii) feature extraction (iv) classification. The aim to for preprocessing is to remove differences in various images taken from different modalities. Image quality may also be required to be enhanced for efficient diagnosis. Segmentation algorithms extract the portion of the image which will be for our interest for analysis. These portions are called region of interest the system highlights which portion of images have higher degree of suspicion of lesions than a given threshold. Then the system is trained with given datasets and uses enough scalable classifier to give reliable results with desirable accuracy more than 90%.

A typical model of Computer assisted diagnosis in Medical imaging : Medical imaging is an emerging discipline of application of artificial intelligence techniques on medical images which are obtained from various medical devices and this discipline will offer visualization and interpretation mechanisms which will draw conclusions which may be overlooked by radiologists. Medical imaging experts work closely with physician so that every minute detail is taken into consideration while designing diagnostic systems.

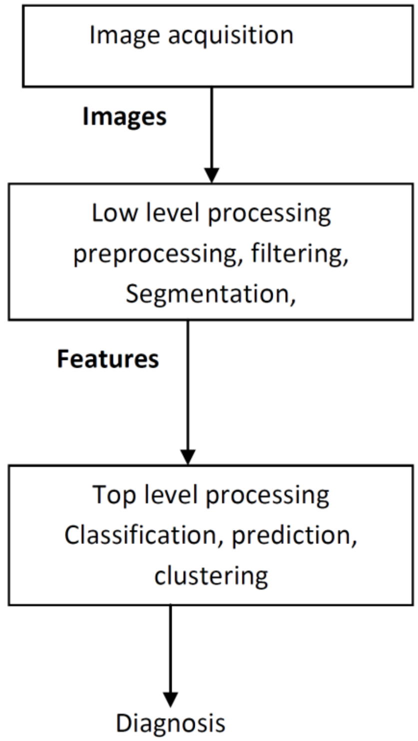

Figure 1. Model of Computer assisted diagnosis in medical imaging systems.

Steps of designing computer assisted diagnosis system: these systems may require interface with hardware and software as well.

- First concerned body organ is scanned by some modality which is recommended by physician, obtained images are in DICOM format.
- Bugs, noise in the images are reduced. Filtering is also performed .It encompasses enhancement, deblurring and edge detection. Image quality is enhanced because if image quality is poor then it would affect the accuracy of classification of disease. Feature extraction and selection helps in obtaining reduced representation of required data. This reduction is such that there is no significant loss of information which may affect the prediction process. Redundant feature of the data is removed and only those features which contribute more for the required task.

- Classification is also known as supervised learning and it works in presence of available training data (labeled data). Classification may two class classification or multi class classification.CAD may classify whether the lesionis benign or malignant. Clustering is unsupervised learning and there is no labeled data available. Clustering may help in finding out similar patients with similar body organ condition which will further draw some useful inference and determine the direction of medication.

Table I. Approaches used in Medical imaging.

Task	**Commonly used techniques**	
Preprocessing	histogram transformation, Laplace Operations	
Filtering	Enhancement	Linear, Non linear ,local or global filter or wavelet based filters
	Blur removal	Inverse or weiner filters
	Detection of Edge	Haar transform
Segmentation		Random Forest
		Deep convolutional networks
Feature Selection		Discriminant Analysis, PCA, ICA
Classification		Euclidean distance, Linear discriminant functions, Bayesian Linear classifier k-NN algorithm, SVM
Clustering		Fischer linear discriminant parsing, k-mean

General Framework for CAD:

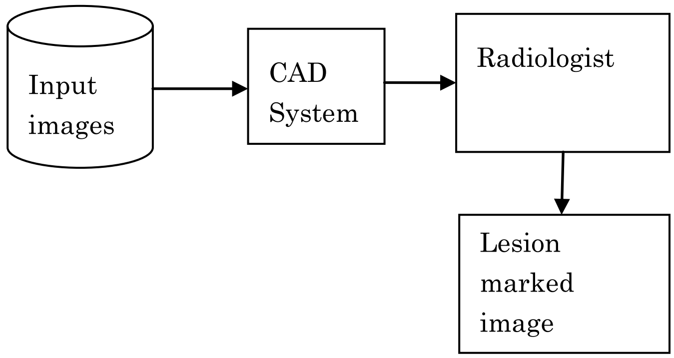

Figure 2. Model of a CAD.

Challenges and performance issues involved in Medical Imaging: The biggest challenging issue with design of CAD system is that it should really reach to the hospitals for actual usage and should not be left lying in computer laboratories. This happens only if it gets approval from competent authority like FDA in USA.FDA approval is given only if it fulfills some guidelines laid down by them. Such systems will be acceptable to radiologists only it increases their efficiency and reduces work load. The main problem with this field is unavailability of too many datasets with which experimentation can be done. One system can be performing well with one particular kind of dataset but may not perform with other dataset. So we may need to experiment with too many dataset of medical images. Creation of such training dataset is a very much time, effort and money consuming. But if such training datasets are available in sufficient numbers, it will help in designing efficient diagnostic system which will be really preferred by radiologists. Computer assisted diagnostic systems analyzes medical images and based on the available training data, it makes decision about the same whether lesion in it is benign or malignant. Overall performance of the system depends upon all sub tasks involved in it. The entire system consists of preprocessing, segmentation, feature extraction and classification. Many techniques have been purported for each of them. No approach is superior to rest others in every case. So we may have to experiment with various techniques of each of them. The training data size also determines the performance of the systems. Larger the training dataset more will be accuracy. Larger training data set consumes a lot of time, effort and cost. Performance of CAD system depends upon number of true-positive and false positive lesionmarkers. The cases which were earlier overlooked by radiologists are now taken care by CAD system. Even if some cased are misclassified by CAD systems can now be taken by radiologists. In a way,CAD systems are going to increase efficiency and reduce workload of radiologists.

Performance Metrics of CAD Systems: Performance of classifier in computer assisted systems is measured by receiver operating characteristic (ROC) curve and confusion matrix. ROC curve plots system's true positive cases versus false positive cases.

True positive (TP) detection rate: percentage of correctly diagnosed cases (presence of disease).

False positive (FP) detection rate: percentage of cases which were not having disease but system diagnosed them as positive.

True Negative (TN) detection rate: percentage of correctly diagnosed cases (absence of disease).

False Negative (FN) detection rate: percentage of cases which were having disease but system diagnosed them as negative.

False positive and false negative both indicate misclassification error. FP+FN is to be reduced in order to design an optimal CAD system.

TP rate indicates how efficient the system is in finding out the presence of the disease. Apart from TP, TN, FP and FN there are two other parameters which determine the performance of a CAD system.

Sensitivity is the probability that patient will be diagnosed positively when he was actually suffering with disease.

$$S_e = \frac{TP}{FN + TP}.$$

Specificity S_p is the metric which indicates that patient will be diagnosed negatively, whereas he actually had not disease.

$$S_p = \frac{TN}{TN + FP}.$$

$$\text{Diagnostic acceuracy} = \frac{TP + TN}{TP + FP + TN + FN}.$$

Area under the curve (AUC) is another metric of comparing classifier performance. For an ideal classifier TP rate is 1 and FP is 0 i.e AUC 0.5.

For a system with better prediction accuracy both S_p, S_e should be high. Training time and actual prediction time should be reasonably low.

Literature Review: in this section we have presented systematic description of most recent work carried out in this field.

In [1], Kupinski and Giger had given a approach to detect breast cancer by mammography images. To train their classifier, they used two techniques :early stopping technique and regularization techniques. Regularization techniques ensures minimization of over training. To reduce training time and increased accuracy, cross validation was utilized in both the techniques. Using these two techniques they obtained an accuracy of 95% and 93% respectively for early stopping and regularization techniques.

In [2], the authors have investigated for detecting micro calcification in Mammographic images to detect breast cancer .They used higher order statistics parameters in classification other preprocessing tasks .They used 3rd and 4th order correlation to detect micro calcification. Mammographic images were preprocessed by subband decomposition filterbank. The partitioning of the bandpass subimage was done into coinciding square regions where skewness and kurtosis are estimated. Critical regions with micro calcification were identified on the basis of high skewness and kurtosis. They experimented this approach on Nijmegen database and obtained that no micro calcification cases were missed.

In [3], William E. Polakowski, Donald A. Cournoyer employed model-based vision (MBV)algorithm to classify lesions into benign or malignant MBV algorithm was proposed to be of five sub tasks to structurally identify critical ROI's, remove false positives, and classify the remaining as benign or malignant. In order to find best features for classification, derivative based feature saliency techniques was used. Nine features are chosen to define the malignant/benign models. The feature extraction module obtains these features from all critical ROI's. This system showed 92% sensitivity.

In [4], Du-Yih Tsai has given Neural network based approach for computer assisted techniques to detect heart disease analyzing ultrasound images. To train the system he used back propagation algorithm. Since weights chosen play crucial role so in order to get optimal weights he used genetic algorithm for it.

In [5], Kenji Suzuki, Junji Shiraishi, have used ANN based classifier for a diagnostic scheme to detect Nodules in Chest Radiographs. Their model consisted of four steps (i) pre-processing based on a difference-image technique. (ii) detection of initial nodule candidates using a multiple gray-level thresholding technique, (iii) grouping of initial nodule candidates, and (iv) application of rule based and linear-discriminant classifiers for reducing false positive results. Difference-image technique is a technique for enhancing lung nodules and suppressing normal background structures. The difference image was obtained by subtraction of the nodule-suppressed image from the nodule-enhanced image. Initial nodule candidates were identified in the difference images through the use of a multiple graylevel thresholding technique. They used supervised nonlinear image-processing techniques based on ANN and applied it for reducing false-positive results in computerized detection of lung nodules.Their scheme removed 68.3% of false-positive cases and sensitivity of 81.3%.

In [6], Ming Li and Zhi-Hua Zhou has presented an approach for designing computer assisted diagnosis which used semi supervised learning which used both labeled and unlabeled data. First classifier was trained using labeled data and its performance by using unlabeled data. These diagnosed samples (labeled data) were prepared by radiologists and physicians. This was expensive step which consumed a lot of time and effort. The Effectiveness of the system was observed on a benchmark data available UCI repository. They investigated the system for diseases like breast cancer. They also investigated that the use of unlabeled data gave a rise of performance by 13.1%.

In [7], R. Bharat Rao, Jinbo Bi designed a system Lung CAD which employed machine learning to detect lung cancer from chest radiographs.

Their CAD system consisted of five stages 1. lung segmentation: to identify the lung area within the chest; 2.Candidate generation which identifies critical suspicious affected candidate regions of interest (ROI) from a medical image; 3. Feature extraction that computes descriptive features for each candidate so that each candidate is represented by a vector x of features; 4.classification that differentiates candidates based on candidate feature vectors;5. Data visualization of CAD findings to the radiologist in order for him to accept or reject the CAD findings. Their system showed a good accuracy that is why FDA gave approval to their CAD system that this can be utilized by radiologists to detect various types of cancer.FDA gives approval to any CAD system for diagnosis only after rigorous clinical trials.

In [8], Adrien Depeursinge, Daniel Sage proposed a method that used a texture classification system that detects lung tissue patterns from high–resolution computed tomography (HRCT) images of people affected with interstitial lung diseases (ILD).It employed Wavelet frames and grey–level histograms & specific features for feature extraction and selection method followed by classification using k-NN method into five lung tissuepatterns (healthy, emphysema, ground glass, fibrosis and microdules. The approach delivers overall multiclass accuracy reaches 92.5%.

In [9], Noah Lee, Andrew F. Laine talked about the capabilities of computer assisted diagnostic systems to detect disease like lung cancer. The purpose was to conduct lung nodule screening and classify into nodule tissue and abnormal lung tissue. Lung nodule categorization consists of two parts (a) nodule detection (b) nodule classification. He argued that apart from detecting cases which overlooked by radiologists, CAD system should also detect any abnormal structure developing in some body organ at an initial stage so that this person could be saved.

In [10], J. Dheeba, N. Albert Singh, S. Tamil Selvi have suggested an approach or detection of breast lesions from mammographic images. In their approach, in classification step, they employed particle

Swarm Optimized Wavelet Neural Network (PSOWNN) and for feature extraction they used Laws Texture features. The final outcome of the system was whether tissue is normal or abnormal (affected with disease). They analyzed the performance of the system using Receiver Operating

Characteristic (ROC) curve. They found out that area under the ROC curve of the proposed algorithm is 0.96853 with a sensitivity 94.167% of and specificity of 92.105%.

In [11], Volodymyr Ponomaryov has given an approach for CAD for breast cancer detection from mammographic images. He used K-Means method for segmentation of preprocessed mammogram images followed by PCA for feature selection method. For classification of breast lesion to benign or malignant he used support vector machine (SVM) and then performance was analyzed by ROC curve. The system obtained accuracy of 94.40%.

In [12], Hoo-Chang Shin and Holger R. Roth applied deep convolution neural networks to computer-assisted detection specially twodiseases thoraco-abdominal lymph node (LN) detection and interstitial lung disease (ILD). They applied deep convolution neural network for image segmentation, training and classification. Their architecture used large number of parameters and large number of layers. They used Image Net which is large dataset of images. Transfer learning was also used to train their system. They obtained 86% sensitivity on 3 false positives (FP) per patient.

In [13], Benjamin Q. Huynh, Hui Li, Maryellen L. Giger applied transfer learning on image features obtained from using CNN on images obtained mammographic digital images to detect early signs of breast cancer. They compared the performance by applying SVM on features extracted by CNN on mammographic images They applied ensemble classifiers to combine prediction of various others classifiers to produce better results. They obtained accuracy of AUC=.86, where AUC is the area under the ROC curve.

In [14], the authors have suggested a deep learning based method to perform segmentation on brain images obtained from MRI scan which can further be used by some CAD system which will this tumor as benign or malignant. They gave an approach based on CNN architecture which is different from those traditionally used in computer vision. CNN utilizes both local features as well as more global features simultaneously. This network utilizes a final layer that is a convolution implementation of a fully connected layer which allows a 40 fold speed up. They have applied a two-phase training procedure that allowed handling problems related to the imbalance of tumor labels. Finally, cascade architecture was explored in which the output of a basic CNN was treated as an additional source of information for a subsequent CNN.

In [15], the authors suggested a model which used textual features which were extracted from the separated ROI's. For classification they used support vector machine. This model candetect whether the input image contains malignant cell or not. Their model gave a precision of 97% for cancer identification 87% for prediction.

Future direction of Research: The images which will be used for this study will be high dimensional. Classical Neural network techniques will not be suitable for such data. Even if they are applied we do not a good level of accuracy. For high dimensional data we can use feature extraction techniques e.g. Principal Component Analysis (PCA) or Independent Component analysis (ICA). For classification we can use high dimensional neural networks which is latest generation of neural networks. High dimensional neural networks can be categorized into complex valued neural networks (CVNN), vector valued neural networks (VVNN) or quaternion valued neural networks (QVNN). High dimensional neural networks has several merits over classical multilayer perceptron networks. For the problem of same complexity, high dimensional neural networks requires smaller network topology (lesser number of learning parameters) and lesser training time and produces much accurate results. So high dimensional neuro computing can be used in computer assisted diagnosis to produce much better results in less amount of time.

References

[1] Matthew A. Kupinski and Maryellen L. Giger, Investigation of Regularized Neural Networks for the Computerized Detection of Mass Lesions in Digital Mammograms, Proceedings-19th International Conference - IEEE/EMBS Oct. 30 - Nov. 2, 1997.

[2] M. NafiGurcan, YaseminYardımcı and A. Enis¸ Cetin, Detection of Microcalcifications in Mammograms Using Higher Order Statistics' IEEE Signal Processing Letters (Volume: 4, Issue: 8, Aug. 1997).

[3] William E. Polakowski, Donald A. Cournoyer and Steven K. Rogers, Computer-Assisted Breast Cancer Detection and Diagnosis of Masses Using Difference of Gaussians and Derivative-Based Feature Saliency' IEEE Transactions on Medical Imaging, Vol. 16, No. 6, December 1997.

[4] Du-Yih Tsai, Classification of heart diseases in ultrasonic images using neural networks trained by genetic algorithms, Fourth International Conference on Signal Processing Proceedings, 1998. ICSP'98.

[5] Kenji Suzuki, Junji Shiraishi, Hiroyuki Abe, Heber MacMahon and Kunio Doi, False-positive Reduction in Computer-assisted Diagnostic Scheme for Detecting Nodules in Chest Radiographs by Means of Massive Training Artificial Neural Network, Academic Radiology, Vol 12, No 2, February 2005.

[6] Ming Li and Zhi-Hua Zhou, Improve Computer-Assisted Diagnosis with Machine Learning Techniques Using Undiagnosed Samples, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans (Volume: 37, Issue: 6, Nov. 2006).

[7] R. Bharat Rao, Jinbo Bi, Nancy Obuchowski and David Naidich, Lung CAD: A Clinically Approved, Machine Learning System for Lung Cancer Detection' KDD '07 Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[8] Adrien Depeursinge, Daniel Sage and Asma Hidki, Lung Tissue Classification Using Wavelet Frames, IEEE Engineering in Medicine and Biology Society, Conference, February 2007.

[9] Noah Lee, Andrew F. Laine, Guillermo Márquez, Jeffrey M. Levsky and John K. Gohagan, Potential of Computer-Assisted Diagnosis to Improve CT Lung Cancer Screening, IEEE Reviews In Biomedical Engineering, VOL. 2, 2009.

[10] J. Dheeba, N. Albert Singh and S. Tamil Selvi, Computer-assisted detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach, Journal of Biomedical Informatics 49 (2014) 45-52.

[11] Volodymyr Ponomaryov, Computer-Assisted Detection System based on PCA/ SVM for Diagnosis of Breast Cancer Lesions, CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), 2015.

[12] Hoo-Chang Shin, Holger R. Roth and Mingchen Gao, Deep Convolutional Neural Networks for Computer-Assisted Detection: CNN Architectures Dataset Characteristics and Transfer Learning, IEEE Transactions on Medical Imaging, DOI 10.1109/TMI.2016.2528162,2016.

[13] Benjamin Q. Huynh, Hui Li and Maryellen L. Giger, Digital mammographic tumor classification using transfer learning from deep convolutional neural networks, Journal of Medical Imaging 3(3), 034501 (Jul.-Sep. 2016).

[14] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, YoshuaBengio, Chris Pal, Pierre-Marc Jodoin, Hugo Larochelle, Brain tumor segmentation with Deep Neural Networks, Medical Image Analysis 35 (2017), 18-31.

[15] Janee Alam and Sabrina Alam Alamgir Hossan, Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifier, International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018.