International Journal of Research in Advent Technology, Special Issue, March 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Computer Aided Identification and Classification of Cancer: A Review

Shivam Chaudhary[1], Nishant Srivastava[2]
[*]Department of Computer Science and Engineering,
Jaypee University Anoopshahr , UP, India
shivasmic07@gmail.com[1]

Abstract— Cancer is one of the most dangerous and life taking disease in the world. However, early diagnosis and treatment can save life. It is difficult for doctors to interpret and identify the cancer from CT scan images. Therefore, Computer vision can play an important role in Medical Image Diagnosis and it has been proved by many existing Computer Aided Diagnosis (CAD) systems. Many computer aided techniques using image processing and machine learning have been researched and implemented. Determining the more efficient methods of detection of cancer to reduce the errors is still a prevailing hurdle among the researchers. The main aim of this research is to evaluate the various computer-aided techniques, analyzing the current best technique and finding out their limitation and drawbacks. Although there are more than hundred types of cancer, the current study is concentrated on Lung cancer, Breast cancer and Brain tumor.

Keywords—lung cancer, breast cancer, brain tumor, computer aided diagnosis, image processing, computer vision, glcm, classification, machine learning

1. Introduction

A malignant growth caused by the uncontrolled division of abnormal cells in any part of the body is known as cancer. There are more than hundred types of cancers. This study will cover three most common types of cancers.

A. Lung Cancer

According to the data made public by World Health Organization (WHO) [1] there were 63475 deaths due to the lung cancer in India in year 2018. Lung cancer is difficult to detect as it arises and shows symptoms in final stage.

B. Breast Cancer

It is the most common malignant disease among the women worldwide. According to the WHO [1] there were *87090 deaths due to this kind of cancer in India.*

C. Brain Tumor

It is a cancerous or non-cancerous growth inside the human brain. There are three kinds of brain tumors i.e, glioma, meningioma and pituitary tumor. According to the WHO [1] there were 24003 deaths in India in 2018.

D. Computer Aided Diagnosis (CAD) System

Imaging plays a very crucial role in the early detection of cancers. MR imaging, CT imaging, digital and analogue Mammography are widely used imaging techniques. But still there lies a challenge for the doctors to efficiently interpret the medical images. Variance of intensity in medical images and misjudgment of anatomical structure by doctors and radiologists might cause difficulty in marking the cancerous cells. Therefore there are relatively high chances of committing errors when these images are examined by the doctors because even though human eyes can perceive a large number of intensity levels but at any particular instant of time they can only differentiate between much smaller numbers. That is when Computer Aided Diagnosis Systems come into the scene. CADs were introduced in early 1980s to assist doctors in interpreting medical images to improve their efficiency and in a very short period of time they became an integral part of cancer diagnosis. Since then a large number of CADs have been designed and implemented. A typical CAD system makes use of a number of digital image processing techniques to improve the quality of the medical image and detect the cancerous tissues. It carries out the extraction of features on the basis of type of cancer it has been searching for. It uses this feature set to train a machine learning model to carry out the classification of the cancer.

Since the invention of CADs a numerous number of modifications have been proposed and implemented into these systems. These modifications involve improving the image preprocessing by implementing the deep learning concepts into the image processing to make preprocessing more efficient and effective [3]. Machine learning is popularly used to carry out the classification hence there is a number of techniques that have been developed and implemented by the researchers in past few years to make the classification process more and more efficient [2]. But still there is a lot to be achieved in terms of the detection and classification efficiency.

The results of this automation in cancer diagnosis are no less than a wonder because according to the annual statistics report from American Cancer Society [4] the death rate from cancer in the US has declined steadily over the past 25 years. As of 2016, the cancer death rate for men and women combined had fallen 27% from its peak in 1991. This decline translates to about 1.5% per year and more than 2.6 million deaths avoided between 1991 and 2016. The drop in cancer mortality is mostly due to steady reductions in smoking and advances in early detection and treatment. Therefore this is evident that the incorporation of Computer Aided Diagnosis systems for cancer has revolutionized the cancer diagnosis. CAD systems have proved their significance in this field since the very beginning.

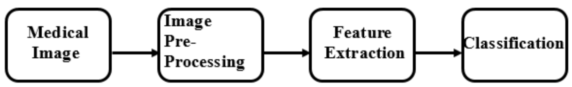

Fig 1. A typical CAD system

Rest of the paper is organized as follows, in section II recent developments that have been proposed by researchers for the Computer Aided Diagnosis of various cancers are discussed. In section III various methodologies which play an important part in Computer Aided diagnosis have been discussed. Finally, the paper is concluded in section IV.

2. LITERATURE REVIEW

Researchers are continuously contributing in the field of Computer Aided Diagnosis of cancer. Several researches have been proposed into the literature by researchers using the concepts of image processing and machine learning. Jin Zhang and Jin [18] made use of convolutional neural networks (CNN) as classifier in their CAD system to detect the lung cancer. Advantage of this model is that it uses circular filter in region of interest (ROI) extraction phase which reduces the cost of training and recognition process. This system has 84.6% of accuracy, 82.5% of sensitivity and 86.7% of specificity. Although there is a reduction in the implementation cost but still their model has unsatisfactory accuracy.

Ignatius and Joseph [19] proposed a CAD system in which they use Gabor filter to enhance the image during preprocessing stage and marker controlled watershed segmentation technique for the detection of cancer nodule. This model extracts the features such as area, perimeter and eccentricity of cancer nodule and shows the accuracy of 91%. This model shows comparatively better results in terms of accuracy but it does not carry out any preprocessing on the cancer nodule neither does it classify the tumor in terms of malignant and benign.

Suren Makaju et al. addresses this issue in the research of Ignatius and Joseph [19] and propose their own version of CAD system [20]. Their proposed CAD system detects the cancer with an accuracy of 92% and guarantees a classification accuracy of 86.6%. However it does not carry out the classification of cancer nodule in terms of stages.
Analyzing the recent developments in the breast cancer diagnosis using CAD systems there is a number of incredible contributions by the researchers. Rahimeh et al. [21] proposed two techniques for segmentation of breast masses using region growing technique. Artificial Neural networks are trained to produce the seeds and thresholds of the segmentation process. The intensity and texture features are extracted and fed into a neural classifier to classify the benign and malignant mammograms. An accuracy of 96.47% has been achieved.

G. Kom et al. [22] proposed a detection algorithm for breast masses inside mammogram which uses a linear transformation enhancement filter for the enhancement of local contrast of each pixel. Local adaptive threshold technique is used for the binarization of the subtracted images from the original image that contains the masses. The sensitivity of the proposed model reaches up to 95.91%.
Punitha et al. [23] in their research work use Gaussian filtering to carry out the preprocessing of medical image. Benign and malignant ROIs containing breast masses are obtained using the Optimized Region Growing technique based on Dragon Fly optimization technique. GLCM and GLRLM features have been extracted after which they are fed to a Feed forward Neural Network. Proposed optimized region growing technique helps the Feed forward Neural Network to achieve an accuracy of 98%.
Recent developments in the field of Computer Aided Diagnosis of brain tumor include the work of Jainy Sachdevaa et al [24]. They have proposed a computer aided diagnosis system to classify and detect tumors. They implemented the Genetic Algorithms (GA) used for optimal feature selection and further the implementation of two hybrid machine learning models GA with support vector machine (SVM) and artificial neural network (ANN) (GASVM and GA-ANN).

El-Sayed Ahmed El-Dahshen et al. [25] proposed a hybrid technique for the classification of brain MR images. It includes three stages namely, feature extraction, dimensionality reduction and classification. They have used Discrete Wavelet Transformations (DWT) for the feature extraction. Two classifiers have been developed for the classification purpose. The first classifier is based on feed forward neural network and second classifier is based on the K-NN (K- Nearest Neighbor) algorithm.

Nelly et al. [26] have carried out a survey based study on brain tumor segmentation. Given the advantages of MRI over other diagnostic imaging techniques, study is focused on the segmentation of MR image of brain. Their study tries to put forward the concepts of semi-automatic and fully automatic segmentation methods along with their advantages and disadvantages.

Eman et al. [27], in their research introduced a new classification technique for the classification of brain tumor MR images. This algorithm is developed by merging two classification methods namely, statistical methods and data comparison methods. Results are obtained by applying the same algorithm on three brain tumor datasets.

3. METHODOLOGIES

A. Imaging

Imaging plays a central role in the diagnosis and management of cancer disease. There is a number of imaging techniques that have been widely used in cancer diagnosis.

These techniques include MRI (Magnetic Resonance Imaging), CT (Computed Tomography), Mammography, Positron Emission Tomography-Computed Tomography (PET/CT), Single Photon Emission Computed Cosmography (SPECT). Although there is a number of imaging techniques available for cancer diagnosis but selection of imaging technique for a particular cancer is still a matter of debate. When it comes to breast cancer a number of imaging techniques have been discussed in literature [5]. Mammography is a technique which is widely used in breast cancer diagnosis. But in recent times digital mammography is rapidly replacing analogue mammography. Breast MRI turns out be an expensive alternative test, with limited availability, but it has the highest sensitivity for cancer of any of the breast imaging techniques. MRI imaging has the potential to provide accurate diagnostic information across a wide range of clinical applications. Ultrasound of the axilla can also be an alternative for breast cancer imaging as the presence of breast cancer metastases within axillary lymph nodes is one of the prognostic symptoms in a patient with breast cancer. MRI and CT scan based imaging remains dominant in lung cancer diagnosis. However a number of modifications have been proposed in these imaging techniques by the researchers [6]. Since whole body MRI provides high tissue contrast for metastasis detection therefore it is useful for the combined detection of local recurrence, regional lymph nodes and distant metastases, and hence it is preferred over CT imaging [6].

When it comes to brain and neck related cancers MRI and CT scan based imaging are preferred. However which one to choose in terms of accuracy is still a controversial topic. Studies carried out by Vidiri et al. [7], and Wiener et al. [8], suggest that MRI offers greater sensitivity and diagnostic efficacy than CT imaging which were found consistent with the studies of X. Qiao et al. [9]. Table 1 shows the imaging techniques for the diagnosis of cancer depending upon the type of cancer.

B. Preprocessing

Early detection of cancer is a promising strategy to cut the mortality rate of cancer. Image Preprocessing is the first and most essential stage of cancer detection which aims to enhance the quality of medical image by removing the irrelevant noises and unwanted parts in the background of medical images. The main purpose of this step is to improve the quality of medical images by removing the unrelated and surplus parts present in the background of cancer images for further processing. Performance of any CAD system heavily depends upon the selection of preprocessing techniques. Image preprocessing mainly involves image enhancement, image restoration and image segmentation.

B. 1. Image Enhancement

It aims to improve the interpretability of an image. It intends to make the image better and enhance it from noise, corruption and interference. It has many components such as image scaling, contrast enhancement, image negative conversion, histogram equalization, etc. But there are few image enhancement techniques which have been popularly used by the researchers while designing their CAD systems.

B. 1. 1. Image Scaling

This technique should be applied on medical images if the images lack standard size of images. Since the medical images may be gathered from different sources and sizes, the first step must be to resize the images to have the fixed pixel width.

B. 1. 2. Contrast Enhancement

It is a beneficial step to improve the interpretability of medical image for further processing. It can sharpen the image border and improve the accuracy by accentuating the brightness difference between background and foreground. Linear contrast enhancement and Non-linear contrast enhancement are two widely practiced techniques in CADs.

- Linear contrast enhancement techniques refers to contrast stretching techniques. The image can be transformed to higher contrast by remapping or stretching the gray-level values so that histogram spread over full range.
- Non-linear contrast enhancement techniques deal with the histogram equalizations and algorithms. They are also widely used for medical purpose.

B. 2. Image Restoration

It is defined as the procedure to recover the degraded image from the noise and blur. Image degradation may happen because of various defects such as imperfection of imaging system, bad focusing, motion, etc. It is essential to know about noises present in the image to select the most appropriate de-noising algorithm.

B. 2. 1. Restoration from noise

It is extremely difficult to apply an effective de-noising algorithm for different types of noisy images. The essential property of a good image de-noising method is to suppress the noise in an image while preserving the edges [10]. There is a number of existing methods for removing the noise and blur from an image. Methods to remove the noise from the image can be classified as Spatial filtering and Transform domain filtering. Spatial filtering consists filters such as Mean filter, Adaptive filter, Order statistics filter. While Transform domain filtering is based upon wavelet transformations. Wavelet transform is an extended form of Fourier transform which represent the functions by wavelets. However, Adaptive median filter, Mean filter and Gaussian smoothing filters are most widely used for the removal of noise in medical applications [11].

B. 2. 2. Restoration from blur

As mentioned earlier blur occurs due to the bad focusing or motion between object and camera. There exist different techniques for de-blurring such as Lucy-Richardson algorithm technique, Inverse filter, Wiener filter de-blurring technique. However, wiener filter is applied as the most powerful technique for de-blurring operations in medical images. It is powerful because it removes the noise from the image as well [12].

B. 3. Image Segmentation

It is the process of subdividing a digital image into multiple segments. It is the process of converting an image into a binary image. It converts an image into an image without background. It partitions the image into various sub-regions to identify meaningful information. It is a very crucial task in the preprocessing stage of any CAD system as it is used to separate the Region of interest (ROI) from the background of the medical image. It involves methods such as gray level thresholding, edge detection and morphological operations such as erosion and dilation.

C. Feature Extraction

This stage plays a pivotal role in the classification of any cancer into different categories. The feature set has to be discriminative enough so that the classification can be accurately carried out. Image features reveal the existing attributions and characteristics of an image. Hence the features taken for the classification must be recognizable, efficient and discriminative. There is various types of features that can be extracted from an image. Some of most commonly used features for classification are shape features, texture features and color features. However in recent researches combinations of different image features, i.e. relation of color and texture [13], color and shape [14], etc. have been proposed in the literature.

C.1 Shape features

Shape of an object refers to its physical structure. Shape can be represented by boundary, region, moment, etc. These representations can be used for matching shapes, recognizing objects. Some of the most common shape features extracted in medical applications are area, perimeter, major axis length, minor axis length, eccentricity, centroid, etc.

C. 2. Texture features

Texture refers to the structural patterns of surfaces of objects such as wood, grain, sand, grass, and cloth. The term texture generally refers to repetition of basic texture elements called texels. A texel contains several pixels, whose placement could be periodic or random. Some of the most common texture features which are widely used in medical applications are energy, correlation, homogeneity, contrast, entropy, maximum probability.

C. 3. Color features

Color features refer to the characteristics of an image related to the presence of color information, like distribution of colors, dominant colors, etc. Color features extraction methods broadly fall in two categories viz. global methods and local methods. In global methods, feature extraction process involves complete image, including global color histogram, histogram intersection, image bitmap, etc. On the other hand local methods only consider a portion of the image, including local color histogram, color difference histogram, etc. Color features can be useful in diagnosis of skin cancer [17].

D. Classification

It is the last and the most decisive stage of any CAD system. This stage involves a process which aims at the classification of a medical image by measuring similarities between the test image and the database images. From a database of medical images a small portion of images is used to train the classifier while another portion of the database is used as test data on which classification is performed. The former one is known as training dataset while the latter one is known as test dataset. The feature set is prepared by carrying out the feature extraction from the training dataset. This feature set is further used for the training of classifier using an appropriate training algorithm. The classification is carried out on the basis of the features that have been extracted from the training dataset and an appropriate calculation metric.

The accuracy of classification mainly depends upon the training algorithm which is used to train the learning model and the discriminative nature of the feature set. Various training models have been proposed and implemented by the researchers through the course of time. But deep learning models such as CNN (Convolutional Neural Networks), FCN (Fully Convolutional Neural Networks) [16], support vector machine and KNN (K-Nearest Neighbor) classifiers are the most commonly used classification models for classification purpose in medical applications. Various improvisations of these models have also been proposed in literature by the researchers in accordance with the type of cancer they are dealing with [15].

Table 1. Preferred Imaging Techniques

Cancer	Imaging Techniques	Preprocessing Techniques	Features	Classification Techniques
Lung	MRI	Depends upon the condition of image.	Shape, Texture, Color	Deep Learning (CNN, FCN)
Breast	MRI	Depends upon the condition of image.	Shape, Texture	Deep Learning (CNN, FCN), SVM
Brain	Mammography, Breast MRI	Depends upon the condition of image.	Shape, Texture.	Deep Learning (CNN),K-NN, SVM

Table 1. shows the preferred techniques that can be used during the different stages of Computer Aided Diagnosis of Cancer on the basis of the studies carried out on the recent developments in the field.

4. CONCLUSION

Prime objective of any Computer Aided Diagnosis (CAD) system is to accurately and efficiently detect the cancer tissues at an early stage so that proper treatment can be given to the cancer patient in order to save his life. Since the introduction of CAD systems into cancer, numerous developments have been proposed in the literature by the researchers in terms of medical imaging, image processing and machine learning. There are certain factors which have to be taken into account while carrying out the Computer Aided Diagnosis of the cancer. For instance what kind of image preprocessing technique has to be applied on the medical image before it can be further processed. What kind of features have to be extracted from the medical image. Deciding the kind of training model and classifier is another thing which has to be taken care of.

It can be analyzed from the recent developments in this field that more emphasis is being given on achieving the maximum accuracy with minimum cost. However, it is observed that the higher accuracy comes with a higher cost.

If a particular CAD system achieves accuracy more than 90% then it also incurs more cost in terms of time. If it achieves less accuracy then it incurs less cost. However it seems that the feature extraction is indeed the most important stage during the functioning of any CAD system. Selection of the most discriminative features according to the type of cancer still remains a challenge for the researchers. The process of feature extraction is of higher priority because the discriminative power of the feature set directly affects the classification process, which in term affects the accuracy of the whole system. However, it seems a CAD system is yet to be developed which works in a generalized way and achieves maximum accuracy and minimum cost for a variety of cancer.

REFERENCES

[1] http://gco.iarc.fr/today/data/factsheets/populations/356-india-fact-sheets.pdf.

[2] Zilong Hu , Jinshan Tang , Ziming Wang , Kai Zhang , Lin Zhang , Qingling Sun , Deep Learning for Image-based Cancer Detection and Diagnosis - A Survey, Pattern Recognition (2018), doi: 10.1016/j.patcog.2018.05.014.

[3] Andreas Maier, Friedrich-Alexander-University Erlangen-Nuremberg, Pattern Recognition Lab,Martensstr. 3, 91058 Erlangen,Germany.

[4] https://www.cancer.org/latest-news/facts-and-figures-2019.html.

[5] Redman A, et al., Imaging techniques in breast cancer, Surgery (2016), http://dx.doi.org/10.1016/j.mpsur.2015.10.004

[6] deSouza NM, et al., Strategies and technical challenges for imaging oligometastatic disease: Recommendations from the European Organisation for Research and Treatment of Cancer imaging group, European Journal of Cancer (2017), https://doi.org/10.1016/j.ejca.2017.12.012J.

[7] Vidiri A, Guerrisi A, Pellini R, et al. Multi-detector row computed tomography (MDCT) and magnetic resonance imaging (MRI) in the evaluation of the mandibular invasion by squamous cell carcinomas (SCC) of the oral cavity. Correlation with pathological data. J Exp Clin Cancer Res 2010;29:73.

[8] Wiener E, Pautke C, Link TM, Neff A, Kolk A. Comparison of 16-slice MSCT and MRI in the assessment of squamous cell carcinoma of the oral cavity. Eur J Radiol 2006;58(1):113–8.

[9] Xianghe Qiao, Wei Liu "Performance of different imaging techniques in the diagnosis of head and neck cancer mandibular invasion: A systematic review and meta-analysis", https://doi.org/10.1016/j.oraloncology.2018.09.024.

[10] N. Radhika, T. Antony,"Image Denoising Techniques Preserving Edge", ACEEE Int. J. on Information Technology, Vol. 01, No. 02, Sep 2011.

[11] B.Shinde, D.Mhaske, M.Patare, A.R. Dani," Apply Different Filtering Techniques to Remove the Speckle Noise Using Medical Images", International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 1, Jan-Feb 2012.

[12] M. S. Ahmed, K. K. Tahbou," Recursive Wiener Filtering for Image Restoration", IEEE Transactions on acoustics, Speech, and Signal Processing, vol. asp -34, No. 4, 1986.

[13] Yang C.C. and Chan M.C, "Color image retrieval based on textural and chromatic features", IEEE International Conference, vol. 4, pp.922-927, 1999.

[14] Gevers T. and Smeulders A.W.M., "PicToSeek: Combing color and shape invariant features for image retrieval", IEEE Trans. Image Processing, vol. 91, pp. 102-119, Jan 2000.

[15] Stephanie Robertson, Hossein Azizpour, Kevin Smith, Johan Hartman, Digital image analysis in breast pathology –from image processing techniques to artificial intelligence, Translational Research (2017), https://doi.org/10.1016/j.trsl.2017.10.010.

[16] Zilong Hu , Jinshan Tang , Ziming Wang , Kai Zhang , Lin Zhang, Qingling Sun , Deep Learning for Image-based Cancer Detection and Diagnosis Ñ A Survey, Pattern Recognition (2018), doi: 10.1016/j.patcog.2018.05.014

[17] Shivangi Jain, Vandana jagtap, Nitin Pise "Computer aided Melanoma skin cancer detection using Image Processing", *Procedia Computer Science 48 (2015) 735–740, doi: 10. 1016/ j.procs. 2015.04.209* .

[18] Jin, X., Zhang, Y., & Jin, Q. (2016) "Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network." 2016 9Th International Symposium On Computational Intelligence And Design (ISCID). DOI: 10.1109/ISCID.2016.1053.

[19] Ignatious, S., & Joseph, R. (2015) "Computer aided lung cancer detection system." 2015 Global Conference On Communication Technologies (GCCT), DOI: 10.1109/GCCT.2015.7342723.

[20] Suren Makaju, P.W.C. Prasad, Abeer Alsadoon, A. K. Singh, A. Elchouemi, "Lung Cancer Detection using CT Scan Images", 10.1016/j.procs.2017.12.016.

[21] Rouhi Rahimeh, Mehdi Jafari, Kasaei Shohreh, Keshavarzian Peiman. Benign and malignant breast tumors classification based on region growing and CNN segmentation. Expert Syst Appl 2015;42: 990e1002.

[22] Guillaume Kom, Tiedeu Alain, Martin Kom. Automated detection of masses in mammograms by local adaptive thresholding. Comput Biol Med 2007;37:37e48.

[23] S. Punitha, A. Amuthan, K. Suresh Joseph, "Benign and malignant breast cancer segmentation using optimized region growing technique", https://doi.org/10.1016/j.fcij.2018.10.005.

[24] Jainy Sachdeva, Vinod Kumar, Indra Gupta, Niranjan Khandelwal, Chirag Kamal Ahuja (2016) "Segmentation, feature extraction, reduction and classification" ELSEVIER, Applied Soft Computing47(2016)151–167.

[25] El-Sayed Ahmed El-Dahshan, Tamer Hosny, Abdel-Badeeh M. Salem (2010) "Hybrid intelligent techniques for MRI brain images classification" Elsevier, Digital Signal Processing 20 (2010) 433–441

[26] Gordillo N, et al, State of the art survey on MRI brain tumor segmentation, Magn Reson Imaging (2013), http://dx.doi.org/10.1016/j.mri.2013.05.002.

[27] Eman M. Ali, Ahmed F. Seddik, Mohamed H. Haggag, "Real Brain Tumors Datasets Classification using TANNN", International Journal of Computer Applications (0975 – 8887) Volume 146 – No.4, July 2016.