

Medical Image Classification Techniques for Early Cancer Diagnosis: A Comprehensive Study

Neelaja K¹, Anand Kumar M², Minavathi³

Assistant Professor, Dept of Computer Science & Engineering, NIE-Institute of Technology, Mysuru, India¹

Research Associate, SERC, Indian Institute of Science, Bengaluru, India²

Professor, Dept. of Information Science & Engineering, PES College of Engineering, Mandya, India³

Abstract: Medical imaging has become a vital part in the early detection, diagnosis and treatment of cancer. The reason: the remarkable power of medical imaging in providing physicians with sight and insight about human disease and physiology. In this role, imaging is used to: Screen, diagnose, and stage cancer; Guide cancer treatments; Determine if a treatment is working; Monitor cancer recurrence; and Facilitate medical research, particularly in such critical areas as drug discovery and therapeutic innovation to improve patient care. The cancer (lung, breast, or brain) biological database which contains the medical images classifies the digital films into three categories: normal, benign and malignant. The normal ones are those characterizing a healthy patient (non-nodules); nodules can be either benign (non-cancerous) or malignant (cancer). This paper presents a comprehensive study of various medical image processing and classification techniques used for predicting various types of cancer in early stage.

Keywords: Computer Aided Diagnosis Systems, Image Mining, Image Recognition, Neural Networks, Classification, Image matching, Image retrieval.

I. INTRODUCTION

Medical imaging often detects cancer at its most curable stage and in many cases, when it is least expensive to treat. As such, medical imaging is center-stage in efforts at cancer prevention and early detection. Often this occurs through screening programs to spot cancer early in individuals who otherwise seem perfectly healthy. In many cases, however, imaging is used to detect tumors or other forms of cancer at their early stages when patients go to the doctor with some symptoms. Molecular imaging offers unique insights into the human body that enable physicians to personalize patient care. A vast amount of image data is generated in daily life and in various fields like medical, astronomy, sports and all kinds of photographic images. It is still at the experimental stage and growing field of research. Lack of understanding in the research issues of image mining is the obstacle to rapid progress [1].

The paper is organized in four sections. Section II provides medical imaging technologies and general process used for cancer detection. Section III lists advantages of using medical imaging in cancer care for preventive healthcare. Section IV covers literature pertaining to image processing and classification algorithms, and applications of these techniques for cancer diagnosis. Section V covers the conclusion.

II. MEDICAL IMAGING TECHNOLOGIES AND PROCESS

The ubiquitous radiographic film that has been the basis of image management for a decade is being replaced by new digital imaging modalities such as Computed Tomography (CT); Magnetic Resonance (MR); Nuclear Medicine (NM); Ultra Sound (US); Digital radiography (DF); Computed Radiography (CR) using storage phosphor imaging plates or film digitizers; Digital Angiography (DA); MR Spectroscopy (MRS); Electron Emission Radiography (EMR).

The Primary imaging technologies used in cancer health care are:

A. X-rays

Two-dimensional Images produced by X-rays are due to the different absorption rates of different tissues, also called radiographs which are used in cancer diagnosis. For example, chest radiographs are often used for early cancer detection or to see if cancer has spread to the lungs or other areas in the chest.

B. Mammography

Mammograms use X-rays to identify tumors or suspicious areas in the breasts. A key benefit of mammography is its ability to identify changes in the breasts before a patient or physician can feel any suspicious abnormalities. Mammograms can be recorded either on conventional film or digitally, using an electronic digital detector. Early detection through mammography has been a major factor in the dramatic decrease in breast cancer mortality.



C. Ultrasound

Also called sonography is an imaging technique in which high-frequency sound waves that cannot be heard by humans are bounced off tissues and internal organs. Their echoes produce a picture called a sonogram.

D. Computed Tomography

It is an imaging procedure that uses special X-ray equipment to create detailed picture. It uses computer intensive reconstruction techniques to create images of the body from X-rays. These are the same type of X-rays used for plain radiographs. However, a radiograph and a CT scan convey different types of information. A CT scanner makes images that are tomographic (cross-sectional) slices of the body (like slices of bread), and the data from a CT scan can be enhanced in several ways to show internal structures more vividly than on plain radiographs. Thus, a physician could not only tell if a tumor is present, but roughly how deep it is in the body.

E. Magnetic Resonance Imaging (MRI)

It uses radio frequency (RF) waves in the presence of a strong magnetic field that passes through the cavity of the MRI machine where the patient lies. These RF waves are used to get tissues to emit radio waves of their own. Different tissues (including tumors) emit a more or less intense signal based on their chemical makeup, so a picture of the body organs can be reconstructed and displayed on a computer screen. Much like CT scans, MRI can produce three-dimensional images of sections of the body, but MRI is often better than CT scans for distinguishing soft tissues.

The other imaging technologies are: Nuclear imaging, Planar Gamma Imaging, SPECT (Single Photon Emission Computed Tomography) Gamma Imaging, The positron emission tomography (PET) scan. In image mining, the process of Detection using classification is generally done using the steps shown in figure 1.

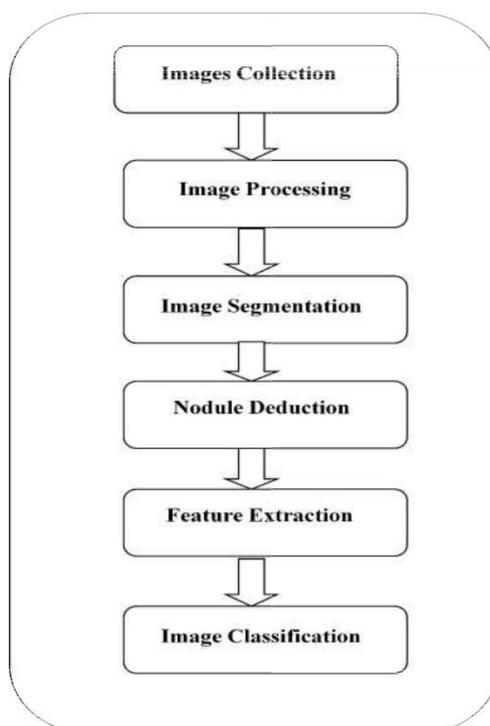


Fig. 1. The general steps of early detection by using data mining method.

III. IMAGING IN CANCER CARE

Imaging Detects Cancer Early: Medical imaging often detects cancer at its most curable stage - and, in many cases, when it is least costly to treat. As medical imaging is center-stage in America's efforts at cancer prevention and early detection. Sometimes this occurs through screening programs to spot cancer early in individuals who otherwise seem perfectly healthy. In many cases, however, imaging is used to detect tumors or other forms of cancer at their early stages when patients go to the doctor with worrisome symptoms. Molecular imaging enables physicians to go even further, detecting disease at the cellular level, before any symptoms or signs are noticeable.

Imaging Enables Less-Invasive Cancer Diagnosis and Treatment: Medical imaging enables a range of less-invasive, highly targeted cancer therapies that translate into better and more comfortable care for patients. Because they



are less invasive, these treatments mean fewer complications, shorter hospital stays, and, in many cases, no incisions or surgery. Thus, the patient benefits in two ways - better care and more comfortable, convenient care.

Mammography Screening Detects Breast Cancer Early: Early detection through mammography has been a major factor in the dramatic decrease in breast cancer mortality. In fact, mammography is one of the key factors in early stage detection-some 63 percent of breast cancers are detected at the localized stage (as a result of mammography screening) for which the 5-year survival rate is 97.5 percent. [2]

Detecting Disease at the Molecular Level: The promise of molecular imaging (i.e., PET imaging), sometimes called the next frontier of diagnostic imaging, is its ability to detect “pre-diseases.” It probes the fundamental biology of disease - that is, changes and abnormalities at the cellular and molecular levels that lead to disease. In most cases, this means identifying the evolution of a disease process, such as cancer, long before signs and symptoms are apparent. [3]

IV. STUDY OF CLASSIFICATION TECHNIQUES

An image retrieval system is a Computer system for Searching, browsing and retrieving images from large databases. Chary et al. [4] described the retrieval of images within a huge collection of images based on color projections and different mathematical approaches which are introduced and applied for retrieval of images. Threshold values are used to subgroup the images, they considered R,G,B color combinations for retrieval of images, which are implemented and results are included, and the results of observation is obtaining efficiency comparatively to the previous one and existing. This method provides the best solution in large image set compared with total of 10000 images with different categories. All suggested methods are prove to perform the good results and based on query images required images retrieved from the database.

P.Rajendran and M. Madheswaran in [5] CT scan brain images are classified using association rule mining. The proposed study categorizes cells into three types namely normal, benign and malign. Low level feature extracted from images and high level knowledge from specialists is combined into system.

Disha Sharma et.al [6] develops an automatic CAD system for early detection of lung cancer by analyzing lung CT images using several steps. First, extracting the lung regions from the CT image using several image processing techniques, this includes bit image slicing, erosion and Weiner filter. Bit plane slicing technique is used in the first step in the extraction process to convert the CT image into a binary image. After extraction, the region growing segmentation algorithm is used to segment the lung region. Then cancer nodules are classified by applying rule based technique. Finally, the extracted features generate a set of diagnosis rules and with the help of diagnostics indicator achieves accuracy of 80%. G Vijay Kumar and Dr GV Raju in [7] proposed early prediction of brain cancer based on texture features and neuro classification logic .The brain tumor in a given MRI image is detected using nine distinct features along with minimum distance. The neuro fuzzy approach is used to recognize the extracted region.

Stanchev [8], using image mining in image retrieval, described a new method for image retrieval using high level semantic features. It is based on extraction of low level color, shape and texture characteristics and their conversion into high level semantic features using fuzzy production rules, derived with the help of an image mining technique. Dempster-hafer theory of evidence is applied to obtain a list of structures containing information for the image high level semantic features. Johannes Itten theory is adopted for acquiring high level color features. The possibility of retrieval using high level image semantic features is the key benefit of this method. The statistic characteristic about the usefulness of the suggested method is obtained after the full system realization.

Dina Aboul Dahab, Samy S. A. Ghoniemy and Gamal M. Selim in [9] applied modified segmentation techniques on MRI scan images to detect brain tumor. Modified Probabilistic Neural Network based on Learning Vector Quantization with image and data analysis to classify brain tumor using MRI scans.

Kailash D.Kharat & Pradyumna Kulkarni [10] proposed two approaches for Brain Tumor classification based on artificial neural networks. The networks were categorized into feed-forward neural networks and back propagation neural Network. First classifier based on feed forward artificial neural network (FF-ANN) and second classifier based on Back propagation Neural Network (BP-ANN). FF-ANN classifier was created with 500 nodes in the first (input) layer. 1 to 50 nodes in the hidden layer and 1 node as the output layer and varied the nodes in order to determine the optimal number of hidden nodes. This was to avoid the fitting or under fitting the data. The most widely used neural network learning method is the BP algorithm. Learning in a neural network involves modifying the weights and biases of the network in order to minimize a cost function. The classifiers have been used to classify subjects as normal or abnormal MRI brain images. In [12], the significant risk factors were identified for particular types of cancer. In this study, a risk factor dataset containing patient records with categorical attributes was first constructed. Then three association rule mining algorithms, Apriori [13], Predictive apriori [14] and Tertius algorithm [15] were employed. The idea of using these algorithms is discovering the most significant risk factors for particular types of cancer to show the highest confidence values. By applying association rule mining algorithms, in addition to finding all combination of the risk factors, the risk factors from frequent risk factors were used to generate the desired rules. Among the three association learning algorithms, the Apriori was the best one to extract useful rules for cancer dataset.



Classification involves the need to find rules that can partition the data into different groups. In health care domain, this type of data mining technique would be important in diagnostic and treatment assistance decision making [16]. The differential diagnosis of erythematous-squamous diseases is a difficult problem in dermatology. They share the clinical features of erythema and scaling with very little differences. The diseases in this group are psoriasis, seboricdermatitis, lichen planus, pityriasis rosea, chronic dermatitis and pityriasis rubra pilaris [17].

There have been several studies reported on erythematous-squamous diseases diagnosis. These studies have applied different methods to the given problem and achieved different classification accuracies. Among these studies the first work on the differential diagnosis of erythematous-squamous diseases was conducted by Guvenir et al. [17]. They developed a new classification algorithm, called VFI5 (for Voting Feature Intervals) and applied it to problem of differential diagnosis of Erythematous Squamous diseases. The VFI5 classification algorithm is an improved version of the early VFI1 algorithm [18]. Guvenir and Emeksiz [19] presented an expert system for differential diagnosis of erythematous-squamous diseases incorporating decision made by three classification algorithms: nearest neighbor classifier [20], naïve Bayesian classifier [21] and voting feature intervals-5 [17]. Nanni [22] developed a new ensemble of support vector machines (SVM) [23] based on Random Subspace (RS) and feature selection. SVM is primarily a dichotomy classifier. It is based on Vapnik–Chervonenkis (VC) dimension and structural risk minimization (SRM), which has been shown to be superior to any other traditional empirical risk minimization principal (ERM). SVM can handle a nonlinear classification efficiently by mapping samples from low dimensional input space into high dimensional feature space with a nonlinear kernel function. In feature space, SVM tries to maximize the generalization problem, and then SVM finds the optimal separating hyperplane, which is also the maximal margin hyperplane. The optimization criterion is the width of the margin between the classes, i.e., the empty area around the decision boundary defined by the distance to the nearest training samples. These patterns, called the support vectors, finally define the classification function. The RS method is the combining technique which modifies the training dataset, builds classifiers on these modified training sets, and then combines them into a final decision rule. In this study, it was obtained the best performance combining the classifiers using the “mean rule” [24].

TABLE I ANALYSIS OF EXISTING TECHNOLOGIES OF IMAGE MINING

Author	Year	Image Pre-processing	Segmentation	Feature Extraction	Classification	Finding Rate
Ed-Edily Mohd. Azhari. Muhd. Mudzakkir Mohd. Hatta. ZawZawHtike and Shoon Lei Win	2014	median filter	Watershed segmentation	-	modified histogram clustering	92%
R.Karuppathal and V.Palanisamy	2014	Speckle Noise Removal Method	FCM	Adaptive histogram equalization method(AHE)	Fuzzy KNN Method	99%
MeghanaNagori. Sivaji Mutkule.rafulSonarkar	2013	graph scanning method	weighted K-means with Z-score algorithm	Graph scanning algorithm	J48graft Algorithm.	94.73%
Ramaswamy Reddy. A.. E.V. Prasad and L.S.S. Reddy	2013	canny edge detection	Fuzzy C-means algorithm	-	-	91%
Pankaj Sapra.Rupinderpal Singh.ShivaniKhurana	2013	Canny's edge detection	modified image segmentation techniques	Automatic utilization of specified regions of interest	modified PNN classifier	100%
Dina AboulDahab.Samy S. A.Ghoniemy.Gamal M.Selim	2012	Canny edge detection algorithm	Region-based segmentation	learning vector quantization (LVQ)	competitive learning neural network(CLN N)	100%
V.P.GladisPushpaRathi and DR.S.Palani	2011	PCA transformation	-	Intensity. Texture. shape based features	LINEAR DISCRIMINANT ANALYSIS	98.87%

Table 1 [11] summarizes various methods which have been used in image processing techniques and also their finding rates varying from 91% to 100%.

V. CONCLUSION

This paper presents the analysis of various data mining and Image processing techniques which can be employed in automated cancer prediction system. Many existing techniques have been employed in recent years for the prediction of different types of cancer at initial stage. This paper presents an overview of various classification and image processing techniques used for predicting cancer. The main focus is on using different classifiers combined with different segmentation algorithms for nodule detection using image processing. The summary of various segmentation and classification techniques with their classification accuracy and sensitivity of nodule detection has been presented. From the study it has been found that for CT scan the most accurate result i.e. 95% is achieved by use of threshold segmentation with Neuro fuzzy classifier, and with PET images higher accuracy of 97% is achieved by use of k-NN and SVM classifier.

ACKNOWLEDGEMENT

The authors of this paper would like to thank researchers and authors/co-authors of all the papers which are referred in preparing this survey paper.

REFERENCES

- [1] T.Y.Gajjar,N.C.Chauhan ,"A review on image mining frameworks and techniques" International journal of computer science and information technologies,vol3(3),2012.
- [2] American Cancer Society, accessed February 10, 2006at http://www.cancer.org/docroot/PED/content/PED_2_3X_maging_Radiology_Tests.asp?sitearea=PED.
- [3] Jelinek, JS et al. "Diagnosis of Primary Bone Tumors with Image-Guided Percutaneous Biopsy: Experience with 110 Tumors." *Radiology*. 223 (2002): 731-737. Also see "Image-Guided Bone Biopsy: Faster, Easier, Safer." RSNA Press Release. May 28, 2002.
- [4] R.Venkata Ramana Chary, Dr.D.Rajya Lakshmi and Dr. K.V.N Sunitha "feature extraction methods for color image similarity" *Advanced Computing: An International Journal (ACIJ)* , Vol.3, No.2, March 2012.
- [5] P.Rajendran and M.Madheswaran, "An Improved Image Mining Technique For Brain Tumour Classification Using Efficient classifier", *International Journal of Computer Science and Information Security*, Vol. 6, No. 3, pp. 107-116, 2009.
- [6] Disha Sharma, Gagandeep Jindal, "Identifying Lung Cancer Using Image Processing Techniques", *International Conference on Computational Techniques and Artificial Intelligence (ICCTAI)*, pp: 115-20, 2011.
- [7] G Vijay Kumar and Dr GV Raju, "Biological Early Brain Cancer Detection Using Artificial Neural Network", *International Journal on Computer Science and Engineering*, Vol. 02, No. 08, pp. 2721-2725, 2010.
- [8] Peter-Stanchev,"Using image mining in image retrieval", *IASTEDconference," Computer science and technology"(2003)*.
- [9] Dina Aboul Dahab, Samy S. A. Ghoniemy and Gamal M. Selim, "Automated Brain Tumor Detection and Identification Using Image Processing and Probabilistic Neural Network Techniques",*International Journal of Image Processing and Visual Communication*, Vol. 1, Issue 2, pp.1-8, October 2012.
- [10] Kailash, D.Kharat,and PradyumnaKulkarni,"Brain Tumor Classification Using Neural Network Based Methods", *International Journal of Computer Science and Informatics*, Vol.1, pp. 2231-5292, 2012.
- [11] A.Sindhu, S.Meera, "A Survey on Detecting Brain Tumor in MRI Images Using Image Processing Techniques", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 3, Issue 1, Jan'15, pp. 123-129.
- [12] J., Tickle, K.S., Nahar, "Significant Cancer Risk Factor Extraction: An Association Rule Discovery Approach," in *Proceedings of International Workshop on Data Mining and Artificial Intelligence*, Khulna, 2008, pp. 108-114.
- [13] R., Imielinski, T., Swami, A.N., Agrawal, "Mining Association Rules between Sets of Items in Large Databases," in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data* , Washington, D.C., 1993, pp. 207-216.
- [14] T. Scheffer, "Finding Association Rules that Trade Support Optimally Against Confidence.," in *Proceedings of the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases(PKDD'01)*, Freiburg, Germany, 2001, pp. 424-435.
- [15] P.A. Flach and N. Lachiche, "Confirmation-guided discovery of firstorder rules with Tertius," *Kluwe Academic Publishers*, pp. 61-95, 2001.
- [16] A.L., Chen, H., Hubbard, S.M., Schatz, B.R., NG, T., Seweel, R.R., Tolle, K.M., Houston, "Medical Data Mining on the Internet: Research on a Cancer Information System," *Artificial Intelligence Review*, pp. 437-466, 2000.
- [17] Guvenir, G. Demiroz, and N. Ilter, "Learning differential diagnosis of Erythematous-Squamous diseases using voting feature intervals," *Artificial Intelligence in Medicine*, vol. 13, pp. 147-165, 1998.
- [18] G. Demiroz and H. A. Guvenir, "Classification by Voting Feature Intervals," in *Proceedings of Ninth European Conference on Machine Learning(ECML-97)*, 1997, pp. 85-92.
- [19] H.A. Guvenir and N. Emeksiz, "An expert system for the differential diagnosis of erythematous-squamous diseases," *Expert Systems with Applications*, vol. 18, pp. 43-49, 2000.
- [20] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms ," *Machine Learning* , vol. 6, pp. 37-66, 1991.
- [21] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*. New York: Wiley, 1973.
- [22] L. Nanni, "An ensemble of classifiers for the diagnosis of erythematous-squamous diseases," *Neurocomputing*, vol. 69, pp. 842-845, 2006.
- [23] T.K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, 1998.
- [24] H.A. Guvenir and N. Emeksiz, "An expert system for the differential diagnosis of erythematous-squamous diseases," *Expert Systems with Applications*, vol. 18, pp. 43-49, 2000.