Automatic Medical Image Classification and Abnormality Detection Using K-Nearest Neighbour

R. J. Ramteke[1], Khachane Monali Y.[2]
School of Computer Sciences, North Maharashtra University, Jalgaon[1,2]

Abstract

This research work presents a method for automatic classification of medical images in two classes Normal and Abnormal based on image features and automatic abnormality detection. Our proposed system consists of four phases Preprocessing, Feature extraction, Classification, and Post processing. Statistical texture feature set is derived from normal and abnormal images. We used the KNN classifier for classifying image. The KNN classifier performance compared with kernel based SVM classifier (Linear and RBF). The confusion matrix computed and result shows that KNN obtain 80% classification rate which is more than SVM classification rate. So we choose KNN algorithm for classification of images. If image classified as abnormal then post processing step applied on the image and abnormal region is highlighted on the image. The system has been tested on the number of real CT scan brain images.

Keywords

K –Nearest Neighbour (KNN), Computed Tomography (CT), Support Vector Machine (SVM), Radial Basis Function (RBF).

1. Introduction

Widely used medical Imaging Computed Tomography modality has been applied in clinical diagnosis. CT assists radiologists to detect and locate Pathological changes with more accuracy. Computed tomography images can be distinguished for different tissues according to their gray level. These images can present information which is constructive for medical diagnosis. The purpose of abnormality detection is to identify the abnormality. Lot of research efforts have been directed towards the field of medical image analysis with the aim to assist in diagnosis and clinical studies.CAD systems are developed to assist radiologists and increase the accuracy of diagnosis.

Medical images are obtained from different modalities like MRI, CT, Ultrasound, PET etc. The CT can be found more reliable technique for early detection of abnormalities like tumor, hemorrhages. This modality provides anatomical information use to plan the direction of radio therapy. For these reason this research work present classification and abnormality detection method for CT brain images.
 K-Nearest Neighbour (K-NN) classification technique is the simplest technique conceptually and computationally that provides good classification accuracy. The k-NN algorithm is based on a distance function and a voting function in k-Nearest Neighbours, the metric employed is the Euclidean distance [1]. SVM have high approximation capability and much faster convergence [2].

2. Literature Review

Padma et.al [3] presented their work for classification and segmentation of CT brain tumor images. Discrete wavelet decomposition and SGLDM method without wavelet transform is used for texture feature extraction.SVM and BPN (Back propagation Network) are used for tumor classification and then tumor is classified from the image.SVM classification accuracy is 96%. Renske de Boer et al [4] was presented work on KNN classifier with additional step for automatic segmentation of white matter lesion in MRI images. Petronella Anbeek et. al [5] proposed A fully automated method for multiple sclerosis (MS) lesion segmentation in T1-weighted MR imaging. Qurat-Ul-Ain et. al [6] proposed approach for Classification and detection of Brain tumor using texture features from brain MR images.SVM used for classification purpose with 99% classification accuracy. Matei Mancas et. al [7] proposed A tumor detection technique on CT scan images of the neck area was proposed based on an airways' symmetry evaluation within slices.

Jose Alex Mathew et. al [8] Different algorithms for threshold detection for abnormality extraction proposed for T1 and T2 MRI Images. IDIS (intelligent Diagnostic Imaging System) GUI was developed by using MATLAB tool. Hassan Najadat [9]

proposed Abnormality detection from CT images of different disease was built based on decision tree classifier that is able to predict general abnormality in human brain. The model was evaluated using hold out method and N-fold evaluation.

Revathy M [1] proposed image classification method to MRI brain imaging using 2nd order wave algorithm for feature extraction and KNN classifier for classification with accuracy 98.6%. Qi Tian, et. al [10] present an automated detection of CT scan slices which contain hemorrhages. Their method is robust towards various rotations, displacement and motion blur. Their work divided into two main parts first Split slices into encephalic region and nasal cavity region. Second detect abnormal encephalic slices.SVM was used for classification.

P.S. Hiremath et. al [11] used the KNN method for Classification of Squamous Cell Carcinoma Based on color and Textural Features in Microscopic Images of Esophagus Tissues. Features set prepared by using Haralick method and mean and standard deviation of images. The texture features are extracted from the luminance channel and the color features are extracted from the chrominance channels. The method applied on images using different color spaces. The experimental results show that the classification accuracy of 100% is obtained using YCbCr color space.

Vanith L. et. al [2] proposed Automatic classification system to classify Ulcer bacteria, Yellow Tuberculosis, red altraistic bacteria, skin bacteria, Plague bacteria and Gut bacteria using Support vector machine. System efficiency was 97.5%. Rajeswari. S et. al [12] MRI brain images are preprocessed using median filter and features are extracted using GLCM and classified into normal, ubnormal-benign and malignant using RBF kernel based multiclass Support Vector Machine. V. Ulagamuthalvi. et. al [13] proposed Ultrasound liver tumor classification into three classes normal, benign and malignant using SVM. Ahmed Kharrat et. al [14] classified Brain MRI images into three classes normal benign or malignant using hybrid approach(SVM + GA) by extracting the features using special gray level dependence method. H. Selvaraj et. al [15] and Khushboo Singh et.al [16] classified Brain MRI slices using linear and nonlinear radial basis function(RBF) kernel based Least Square Support Vector Machine. LS SVM outperforms when compared with other classifiers like SVM with linear,

SVM with RBF, Multilayer Perceptron (MLP) and K-NN. Classification rate is 99.64%. Traditional Chinese Medicine (TCM) tongue diagnosis is one of the important methods.

Y. Ireaneus et. al [17] proposed Mammogram breast cancer images classification methodology with focusing on two main problems are How to detect tumors as suspicious regions with very weak contrast to their background and How to extract features which categories tumors. Mammogram enhancement includes filtering top hat operations; DWT. Contrast stretching is used to increase the contrast of the image. Using thresholding image is segmented. The features are extracted from segmented breast area then region is classified using SVM.88.75% sensitivity achieved for this method. Zhong Gao et. al [18] proposed   Toung inspection method based on chromatic and textural features using multiclass SVM was developed and average 86.6 % classification rate achieved.

3. Materials and Methodology

Image Data
The real data used in this work collected from CT scan centres. The data contain normal as well as abnormal CT scan brain images.51 CT scan normal and abnormal images for training purpose.

Features
Texture is   one of the important characteristics used in identifying objects or regions of interest in an image. Haralick proposed some easily computable texture features based on gray tone spatial dependencies [19].The proposed method based on the cooccurrence matrix and Haralick features that characterize the color and texture of image. This method is classical and in the pattern recognition community and has extensively been used on gray scale images. Let I be a gray scale image coded on m gray levels. Let $s(x,y)$ be the position of a pixel in I and $t(\Delta x, \Delta y)$ be a translation vector. The co occurrence matrix M_t is a m x m matrix whose $(i, j)^{th}$ element is the number of pairs of pixels separated by the translation vector t that have the pair of gray levels (i, j).

$$M(i,j) = card\{(s, s+t) \in R^2 | I(s) = i, I(s+t) = j\} \quad (1)$$

The choice of the relative position vector is the same as Haralick's [10]. The extracted features are

Contrast (C), Homogeneity (H),Correlation (Cor),Entropy (E) and Local Homogeneity(LH).

Contrast(C): This measure analyses the image contrast (locally gray-level variations) as the linear dependency of grey levels of neighbouring pixels (similarity). Typically high, when the scale of local texture is larger than the distance.

$$C = \sum_{k=0}^{m-1} k^2 \sum_{|i-j|=k} M(i,j) \qquad (2)$$

Homogeneity (H):

$$H = \sum_i \sum_j (M(i,j))^2 \qquad (3)$$

Correlation (Cor): This measure analyses the linear dependency of grey levels of neighbouring pixels. Typically high, when the scale of local texture is larger than the *distance*.

$$Cor = \frac{1}{\sigma i \sigma j} \sum_i \sum_j (i - \mu i)(j - \mu j) M(i,j) \qquad (4)$$

Entropy (E): This measure analyses the randomness. It is high when the values of the moving window have similar values. It is low when the values are close to either 0 or 1

$$E = \sum_i \sum_j M(i,j) log(M(i,j)) \qquad (5)$$

Local Homogeneity (LH):

$$LH = \sum_i \sum_j \frac{M(i,j)}{1+(i-j)^2} \qquad (6)$$

Methodology
The proposed method contains multiple phases as shown in Fig.1 and Fig 2 describes complete method in detail.

Phase1: Pre-processing
Pre-processing includes conversion of Input image into gray scale image and noise removal. In this work median filter used for de-noising.

Figure 1: Phases of Proposed Work

In median filtering, the neighbouring pixels are ranked according to brightness (intensity) and the median value becomes the new value becomes the new value for the central pixel.
A median filter offers three advantages
1.   No reduction in contrast across steps since output values available consists only those presents in the neighborhood.

2.   Median filtering does not shift boundaries.
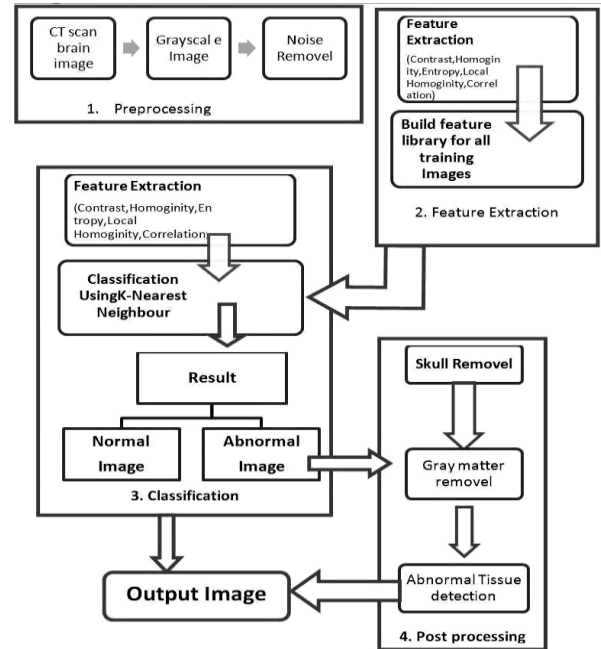3.   The median is, in a sense a more robust average that the mean, as it is not affected by outliers [12].

Figure 2: Proposed Methodology for abnormal classification and detection

Phase 2: Feature Extraction
The purpose of feature extraction is to reduce the original data set by measuring certain properties of features that distinguish on input pattern from another [15].Five statistical features namely Contrast, Entropy, Homogeneity, Correlation, Local Homogeneity are extracted by using Haralick texture feature extraction method. These extracted features used for preparation of Training data. These features are saved into feature library.

Phase 3: Classification
Classification phase classifies the input test image into Normal or Abnormal Class by using K-Nearest Neighbour algorithm. KNN was chosen for classification purpose after verifying its classification accuracy with SVM. Normal Classified image displayed as resultant normal image. Abnormal Classified image passed to next phase for further processing.

•   K-Nearest Neighbour (KNN)
A simple classifier is a NN classifier where each pixel is classified in the same class as the training

data with the closest intensity. There is a possibility of an NN classifier yielding an erroneous decision if the obtained single neighbour is an outlier of some other class. To avoid this and improve the robustness of the approach the KNN classifier works with K patterns .The KNN classifier is considered a non parametric classifier since it makes no underlying assumption about the statistical structure of the data [15].

KNN Algorithm
1. Training set includes classes.
2. Examine K items near item to be classified.
3. New item placed in class with the most number of close items.
4. O(q) for each tuple to be classified. (Here q is the size of the training set.).

The K-nearest neighbour classification is performed by using a training data set which contains both the input and the target variables and then by comparing the test data which contain only the input variables to that reference set the distance of the unknown to K nearest neighbours determines its class assignment by either averaging the class numbers of the K nearest reference points or by obtaining a majority vote for them [20].

- Support Vector Machine (SVM)

SVM is a powerful supervised classifier and accurate learning technique that has been introduced in 1995. SVM has proven its efficiency over neural networks and RBF classifiers.SVM uses an optimum linear separating hyper plane to separate two set of data in feature space. This optimum hyper plane is produced by maximizing minimum margin between the two sets. Therefore the resulting hyper plane will only be dependent on border training patterns called support vectors.SVM operates on two mathematical operations-

- Nonlinear mapping of an input vector into a high dimensional feature space that is hidden from both the input and output.
- Construction of an optimal hyper plane for separating the features [2].

SVM maps input vectors to a higher dimensional vector space where an optimal hyper plane is constructed [12].

The data with linear severability may be analyzed with a hyper plane, and the linearly non separable data are analyzed with kernel functions such as Gaussian RBF.

$$K(x, y) = x^T y \quad \text{(Linear SVM)} \qquad (7)$$

$$K(x, y) = \exp\left(-\frac{\|x-y\|^2}{\sigma^2}\right) \quad \text{(RBF)} \qquad (8)$$

The output of an SVM is a linear combination of the training examples projected onto a high dimensional feature space through the use of kernel function [15]. SVM is based on the structured risk minimization principle from the statistical learning theory. Its kernel is to control the empirical risk and classification capacity in order to minimize the margins between the classes and minimize the true cost [14].

For this work SVM with kernel function linear and RBF (Radial Basis Function) is used for classification of images into two classes namely "Normal" and "Abnormal". The labels for these classes are using '1' and '2' for "Normal" and "Abnormal" respectively. Classification Performance results are discussed in Result section in detail. On the basis of classification accuracy rate KNN algorithm is chosen for the classification purpose.

Phase 4: Post processing
Post processing includes detection of abnormal region from CT input image. The Post processing step remove skull from image then gray matter removed and the abnormal region extracted. The Normal CT image is of 255 gray level image matrix containing values from 0 to 255 for pixel. In CT Brain image the brightest part in the image is skull. So for removing skull we use the brightest pixel values and remove them from the image. The brain matter appears in gray color in the image and if there are certain abnormalities in the brain then it appears bright then the gray matter. So for abnormality detection purpose we used these intensity values.

4. Results

Training data contains 51 images features. Test input image is classified by using KNN classifier. Normal class image is the output image no any further process applied on this image. But abnormal class image processed further for ROI detection. The method is implemented by using MATLAB 2011. Fig 3 shows the ROI detection in abnormal class image.KNN Classifier compared with SVM classifier and the classification accuracy calculated by using confusion matrix. Table 1, Table 2 and Table 3 show the confusion matrix for KNN and SVM (Linear) and SVM (RBF) respectively.
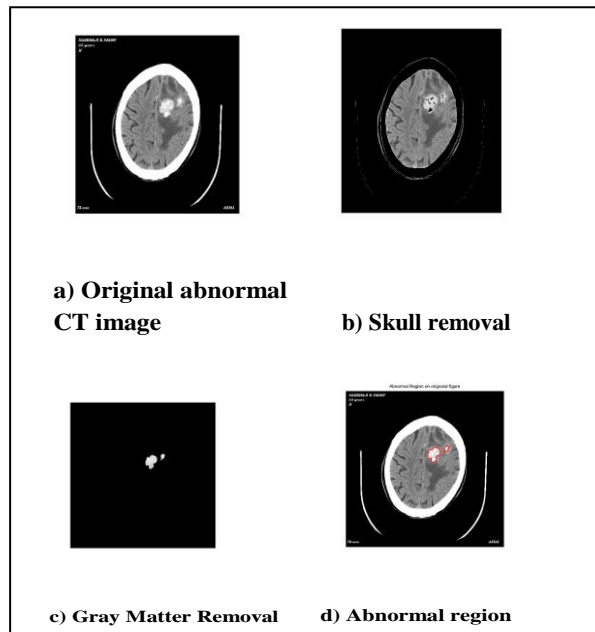
a) Original abnormal CT image   **b) Skull removal**

c) Gray Matter Removal   **d) Abnormal region**

Figure 3: Abnormality Detection Method

Table 1: Confusion Matrix for KNN Classifier

	Normal	Abnormal	Total
Normal	21	5	26
Abnormal	5	20	25

Table 2: Confusion matrix for SVM (Support Vector Machine) classifier with linear kernel

	Normal	Abnormal	Total
Normal	20	6	26
Abnormal	11	14	25

Table 3: Confusion matrix for SVM (Support Vector Machine) classifier with RBF kernel

	Normal	Abnormal	Total
Normal	21	5	26
Abnormal	11	14	25

Using data from Table 1, Table 2, Table 3, we computed sensitivity, specificity, and accuracy for both evaluations. These measures are listed in Table 4.Sensitivity is calculated by using following equations.

$$Sensitivity = \frac{TP}{TP+FN} * 100\% \qquad (9)$$

$$Specificity = \frac{TN}{TN+FN} * 100\% \qquad (10)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \qquad (11)$$

Where
True Positive (TP) - Sick people correctly diagnosed sick, False Positive (FP) - Healthy people incorrectly identified as sick.

True Negative, (TN)-Healthy people correctly identified as healthy. False Negative, (FN)-Sick people incorrectly identified as healthy [12].

Table 4 : Comparison of Classifiers

Evaluation Test / Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)
KNN Classifier	80	81	80
SVM with linear kernel Classifier	77	56	67
SVM with RBF kernel Classifier	80	56	69

The evaluation measurements have shown that KNN classifier shows 80% classification rate.

5. Conclusion

This work proposed automatic classification and abnormality detection from CT scan brain images accurately. The input CT brain image classified as Normal or Abnormal by using KNN classifier with 80 % classification rate. Classified Abnormal image is processed further. Post processing applied on abnormal image and region of interest highlighted on the original image by proposed work. This method will be extended for abnormality detection from multimodal brain images and improving classification rate.

References

[1] Revathy M., "Image Classification with Application to MRI Brain using 2nd Order Moment Based Algorithm", International Journal of Engineering Research and Applications (IJERA) ,Vol.2 Issue 3, pp.1821-1824, May-Jun 2012.

[2]   Vanitha. L. and Venmathi. A.R, "Classification of Medical Images Using Support Vector Machine", International Conference on Information and Network Technology IACSIT Press, Singapore,vol.4,pp 63-67,2011.

[3]   A. Padma and Dr.R. Sukanesh ,"Automatic Diagnosis of Abnormal Tumor Region From Brain Computed Tomography Images Using Wavelet Based Statistical Texture Features" , International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.1, No.3, August 2011.

[4]   Renske de Boer, Fedde van der Lijn, Henri A. Vrooman,Meike W. Vernooij,M. Arfan Ikram, Monique M.B. Breteler, Wiro J. Niessen ,"Automatic segmentation of brain tissue and white matter lesions in MRI", Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium, pp. 652 – 655.

[5]   Petronella Anbeek, Koen L. Vincken and Max A. Viergever, "Automated MS-Lesion Segmentation by knearest Neighbor Classification", Insight Journals(United State),14 July 2008.

[6]   Qurat-Ul-Ain, Ghazanfar Latif, Sidra Batool Kazmi, M. Arfan Jaffar, Anwar M. Mirza , "Classification and Segmentation of Brain Tumor using Texture Analysis", Recent Advances In Artificial Intelligence, Knowledge Engineering And Data Bases, pp 147-155 .

[7]   Matei Mancas, Bernard Gosselin, and Benoît Macq, "Tumor Detection using Airways Asymmetry", Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005, 27th Annual International Conference, 2005.

[8]   Jose Alex Mathew, A. M. Khan, U. C. Niranjan ,"Algorithms to find the thresholds for the Abnormality Extraction of the MRI slice Images of a GUI based Intelligent Diagnostic Imaging System", International Conference on VLSI, Communication & Instrumentation (ICVCI), Proceedings published by International Journal of Computer Applications (IJCA), pp.18-23,2011.

[9]   Hassan Najadat, Yasser Jaffal, Omar Darwish, Niveen Yasser, "A Classifier to Detect Abnormality in CT Brain Images",Proceeding of the International Multiconference of Engineers and Computer Scientists 2011 , Hong Kong Vol1 I, March 16-18,2011.

[10] Ruizhe Liu, Chew Lim Tan, Tze Yun Leong , Cheng Kiang Lee, Boon Chuan Pang, C. C. Tchoyoson Lim, Qi
Tian, Suisheng Tang, Zhuo Zhang ,"Hemorrhage Slices Detection in Brain CT Images", IEEE 978-1-4244-2175-6/08 ,2008 .

[11] P.S. Hiremath, Humnabad Iranna Y., and Jagadeesh D. Pujari,   "Classification of Squamous Cell Carcinoma Based On Color and Textural Features In Microscopic Images of Esophagus Tissues", Journal of Computer Science 3 (7),
pp 566-573, 2007 .

[12] Rajeswari. S , Theiva Jeyaselvi. K, "Support Vector Machine Classification For MRI Images", International Emerging Trends in Computer and Electronics Engineering Dubai,Journal of Electronics and Computer Science Engineering Volume1, Number 3, March 24-25, 2012.

[13] V. Ulagamuthalvi, D. Sridharan; "Automatic Identification of Ultrasound Liver Cancer Tumor Using Support Vector Machine"; International Conference on Emerging Trends in Computer and Electronics Engineering, pp 41-43.

[14] Ahmed Kharrat, Karim Gasmi, Mohamed Ben Messaoud, Nacéra Benamrane And Mohamed Abid; "A Hybrid Approach for Automatic Classification of Brain MRI Using Genetic Algorithm and Support Vector Machine"; Leonardo Journal of Sciences ISSN 1583-0233 Issue 17, July-December 2010.

[15] H. Selvaraj, S. Thamarai Selvi, D. Selvathi, L. Gewali, "Brain MRI Slices Classification Using Least Squares Support Vector Machine", *IC-MED ,Vol. 1, No. 1, Issue 1,* pp21-33 ,2007.

[16] Khushboo Singh, Satya Verma ; "Detecting Brain Mri Anomalies By Using Svm Classification", International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 4, pp.724-726, June-July 2012.

[17] Y.Ireaneus Anna Rejani, Dr.S.Thamarai Selvi; "Early Detection Of Breast Cancer Using SVM Classifier Technique",International Journal on Computer Science and Engineering, Vol.1 (3), pp. 127-130, 2009.

[18] Zhong Gao, Lai-Man Po, Wu Jiang, Xin Zhao and Hao Dong; "A Novel Computerized Method Based on Support Vector Machine for Tongue Diagnosis", In proceeding of: Signal-Image Technologies and Internet-Based System, ISBN: 978-0-7695-3122-9 ,2007.

[19] Robert   M.Haralick,K.   Shanmugam,It'shak Distein , "Textural Features for Image Classification" ,IEEE Transactions on systems, Man and Cybernetics, Vol 1973.

[20] Hiremath P.S., Parashuram Bannigidad, Manjunath Hiremath, "Digital Microscopic Image Analysis of Virus Particles", International Journal of Machine Intelligence, Issue 4, pp 180-184, 2011.

Dr. R. J .Ramteke MSc.(computer Science),SET, Ph.D Associate Professor, Head, Department of Information Technology, North Maharashtra University, Jalgaon.

Khachane Monali Yashawant (04-10-83) Research Scholar, M.Sc.(computer Science),SET, Ph.D (appeared), North Maharashtra University,Jalgaon.