ISSN XXXX XXXX © 2018 IJESC

Research Article Volume 8 Issue No.5

Brain Tumor Detection Using Principle Component Analysis and Support Vector Machine

Anand M. Patil[1], Aneesa S. Nagarakatti [2], Pankaja B. Chichagandi[3], Megha V[4], Prof.Sagar Birje[5]
BE Student[1, 2, 3, 4], Assistant Professor & HOD[5]
Department of Computer Science and Engineering
Angadi Institute of Technology and Management, Belagavi, India

Abstract:
Automated and accurate classification of MRI brain images is extremely important for medical analysis and interpretation. Over the last decade, numerous methods have already been proposed. In our project, we presented a novel method to classify a given MR images as normal or abnormal. The proposed method first employed the wavelet transform to extract features from images. The element extraction is a method of representation of the image with the raw data by performing the processing to extract the useful data from the image to improve the process of decision-making like the classification of different patterns. Followed by applying principal component analysis(PCA) to reduce the dimensions of the features. The reduce features were submitted to a kernel support vector machine (KSVM). This method combines the intensity and the components of the shapes and different orders with that texture of the tumour from the MRI images. SVM is the automatically classified brain MRI images under two categories, either normal or abnormal. The confirmation of normal and abnormal MRI images is based on consistency which is manifest clearly in the axial and coronal images. Using feature vector gained from the MRI images. SVM classifier is used to classify the images. We choose common brain diseases as abnormal brains, and collected 20 MR brain images from Harvard Medical school website. We performed our proposed method with four different kernels and found that linear kernel is one of the best classifications which gives a more accuracy.

Keywords: Feature Extraction, Principal Component Analysis (PCA), Support Vector Machine (SVM), Discrete Wavelet Transform (DWT), Feature Selection

I. INTRODUCTION

Automatic Magnetic resonance imaging (MRI) is a non-invasive medical imaging technique used in Computer Aided diagnosis (CAD) to visualize detailed internal structure and limited functions of the body such as brain disease, Alzheimer disease or movement’s disorders such as Parkinson. The diagnostic value of MRI is greatly magnified by the automatic and accurate classification of the MR images. Different algorithms are disclose in each step on automatic brain tumour detection, this process has its suitable routine. One of the most powerful methods for extraction is Wavelet transform. This is an effective tool for 2D image feature extraction because it allows for analysis of images at the various level of resolution. The main use of the wavelet transform is that it provides localized frequency information about the function of the signal, which is particularly advantageous for classification. Wavelet is applied for reduction of samples and removes high-frequency noises. However, it requires large storage and is computationally more expensive. Hence another routine for dimension reduction scheme is used. In order to reduce the feature vector dimensions and increase the discriminative power, the principal component analysis (PCA) is used. PCA is appealing since it effectively reduces the dimensionality of the data and therefore reduces the computational cost of analyzing new data. After obtaining the feature set, we need to construct the classifier. Also at classification step different algorithm are presented. The first category is unsupervised classification; the other category is supervised classification such as support vector machine (SVM) that classifies points by allowing to one of two disjoint half spaces. These half spaces are either in the original input space of the problem for linear classifiers or in a higher dimensional feature space for nonlinear classifiers. SVM is used for classification as it gives better accuracy and performance than other classifiers. In this project, the proposed method is divided into five stages. The stages are, feature extraction by DWT, feature selection by PCA, feature reduction by k-means algorithm. The output of this phase which was a selected set of features is used as entries to the classification step. Classification is done using SVM.SVM classification distinguishes brain abnormality according to its training.

II. OBJECTIVES

- To extract features from images
- To apply principles component analysis (PCA) to reduce the dimensions of features.
- To apply reduced features to a kernel support vector machine (KSVM).
- To classify a given MR brain image as normal or abnormal

III. EXISTING SYSTEM

- In this existing system one of the most powerful methods for extraction is Wavelet transform. This is an effective tool for 2D image feature extraction because it allows for analysis of images at the various level of resolution.
- Wavelet is applied for reduction of samples and removes high-frequency noises.

DISADVANTAGES

- It requires large storage.

- It is computationally more expensive.

IV. PROPOSED SYSTEM

- In this proposed system we used the PCA as a tool for transforming the existing input features into a new lower dimension feature space.
- PCA is used for the data extraction from the images.
- SVM is used for classification as it gives better accuracy and performance than other classifiers.

ADVANTAGES

- It effectively reduces the dimensionality of the data.
- It reduces the computational cost of analyzing new data.

V. WORKING

In total, our method consists of three stages:

Step 1. Pre-processing (including feature extraction and feature reduction)

Step 2. Training the kernel SVM

Step 3. Submit new MRI brains to the trained kernel SVM, and output the prediction.

We will explain the detailed procedures of the pre-processing in the following subsections.

A. Feature Extraction

The most conventional tool of signal analysis is Fourier transform (FT), which breaks down a time domain signal into constituent sinusoids of different frequencies, thus, transforming the signal from time domain to frequency domain. However, FT has a serious drawback as discarding the time information of the signal. For example, analyst cannot tell when a particular event took place from a Fourier spectrum. Thus, the quality of the classification decreases as time information is lost. It provides some information about both time and frequency domain. However, the precision of the information is limited by the size of the window. Wavelet transform (WT) represents the next logical step a windowing technique with variable size. Thus, it preserves both time and frequency information of the signal. Another advantage of WT is that it adopts scale" instead of traditional frequency", namely, it does not produce a time-frequency but a time-scale view of the signal. The time-scale view is a di®erent way to view data, but it is a more natural and powerful way, because compared to frequency", scale" is commonly used in daily life. Meanwhile, in large small scale" is easily understood than in high low frequency".

B. 2D DWT

Using 2D Wavelet transform (Daubechies-two of level one), the brain image is decomposed into four sub-bands. The clearest appearance of the changes between the different textures represented by the sub-band whose histogram has the maximum variance.

Daubechies-two level 1 wavelet approximation coefficients of the MR brain images are extracted from this sub-band. In case of 2D images, the DWT is applied to each dimension separately. Fig 4illustrates the schematic diagram of 2D DWT. As a result, there are 4 sub-band (LL, LH, HH, and HL) images at each scale. The sub-band LL is used for next 2D DWT. the image, while the LH, HL, and HH sub bands can be regarded as the detailed components of the image. As the level of decomposition increased, compacter but coarser approximation component was obtained. Thus, wavelets provide a simple hierarchical framework for interpreting the image information. In our algorithm, level-3 wavelet decomposition tree was utilized to extract features. The border distortion is a technique issue related to digital alters which is commonly used in the DWT. As we alter the image, the mask will extend beyond the image at the edges, so the solution is to pad the pixels outside the images. In our algorithm, symmetric padding method was utilized to calculate the boundary value.

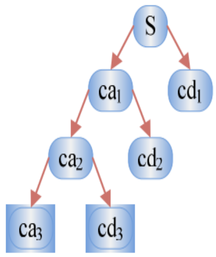

Figure.1. A 3-level wavelet Decomposition tree.

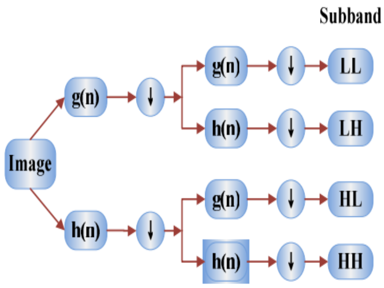

Figure.2. Schematic Diagram of 2D DWT

The LL sub band can be regarded as the approximation component of the image, while the LH, HL, and HH sub bands can be regarded as the detailed components of the image. As the level of decomposition increased, compacter but coarser approximation component was obtained. Thus, wavelets provide a simple hierarchical framework for interpreting the image information. In our algorithm, level-3 decomposition via Harr wavelet was utilized to extract features. The border distortion is a technique issue related to digital filter which is commonly used in the DWT. As we filter the image, the mask will extend beyond the image at the edges, so the solution is to pad the pixels outside the images. In our algorithm, symmetric padding method was utilized to calculate the boundary value.

C. Feature Reduction

Excessive features increase computation times and storage memory. Furthermore, they sometimes make classification more complicated, which is called the curse of dimensionality. It is required to reduce the number of features. PCA is an efficient tool to reduce the dimension of a data set consisting of a large number of interrelated variables while retaining most of the variations.

It is obtained by transforming the data set into a new set of ordered variables according to their variances or importance. This technique has three effects: it orthogonalizes the components of the input vectors so that uncorrelated with each other, it orders the resulting orthogonal components so that those with the largest variation come first, and eliminates those components contributing the least to the variation in the data set. It should be noted that the input vectors be normalized to Progress In Electromagnetics Research, Vol. 130, 2012 375 have zero mean and unity variance before performing PCA. The normalization is a standard procedure.

VI. ARCHITECTURAL DIAGRAM AND SEQUENCE DIAGRAM

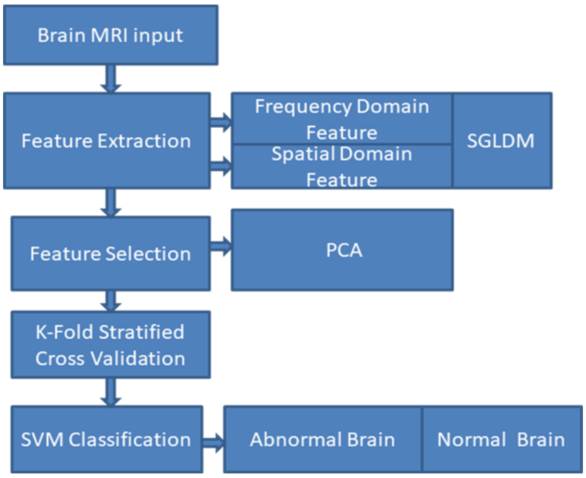

Figure.3. Architectural Diagram.

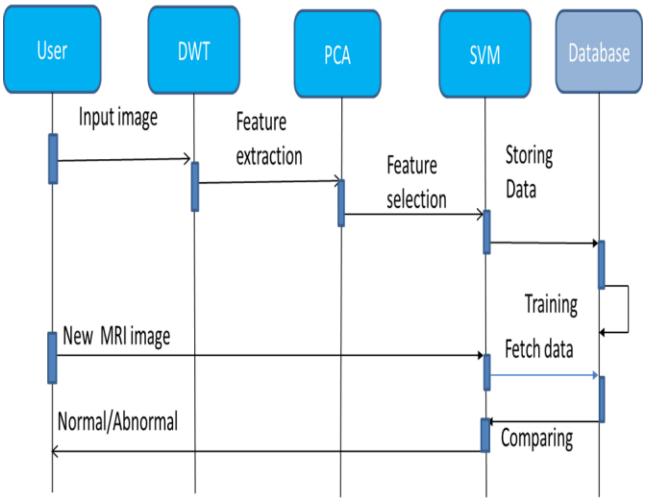

Figure.4. Sequence Diagram.

VII. RESULTS

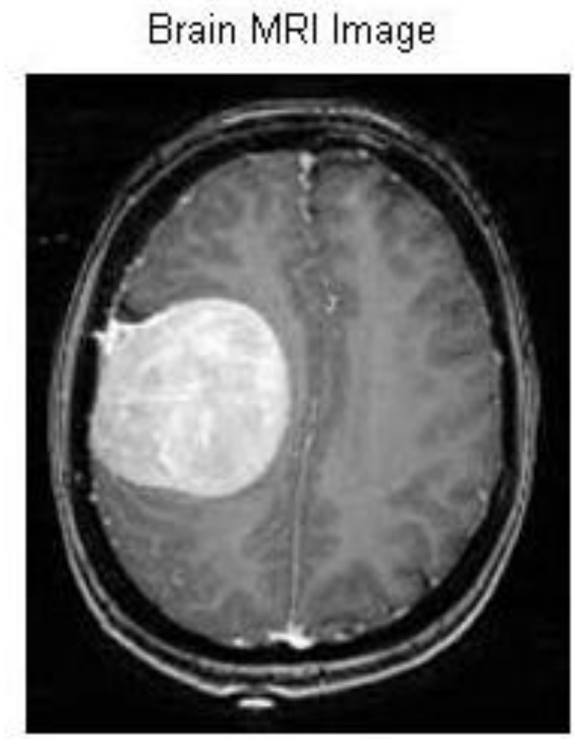

Figure.4. Sample input image for classification

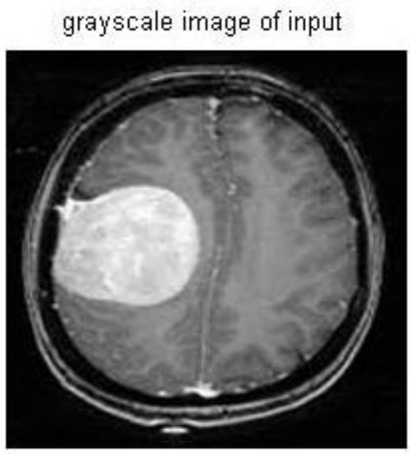

Figure.5. greyscale version of input image

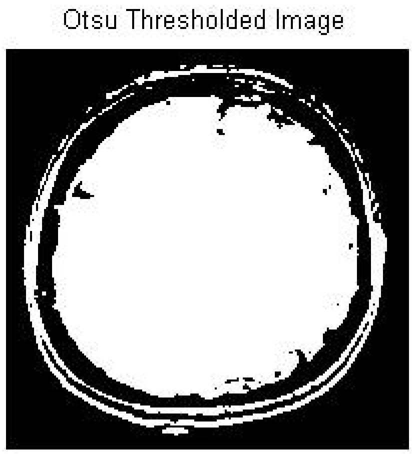

Figure.6. greyscale version of input image

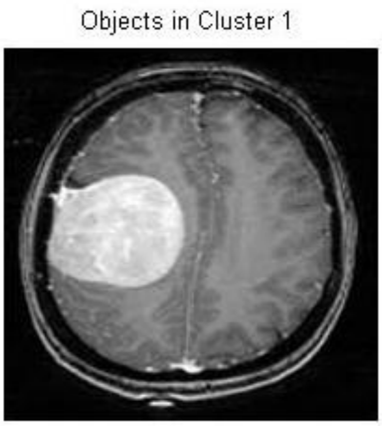

Figure.7. ROI of input image

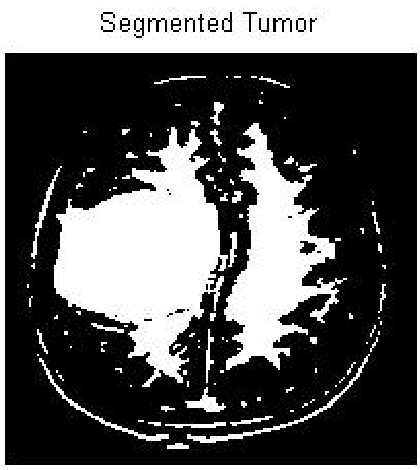

Figure.8. Binary format of ROI image

To have a closer look at the results and interim outcomes of this work varieties of inputs are necessary. Some of the variations in different brain images are to be observed.

In figure 4 the input is a typical benign tumour image. The processing steps and interim outputs of each one of them are to follow. **In figure 5** the input image is first converted into grey format representation. **In figure 6** the boundary identification is as shown i.e. the black and white or binary version of the input image. **In figure 7** Image input is segmented using k means. The outcomes will be as many as there are choices however the region of interest is the one lowest centered**. In figure 8** Taking this as input the binary conversion of image is as shown. The features obtained after the segmentation process are used for SVM. The typical features set for benign input image is:
feat = [Contrast ,Correlation, Energy, Homogeneity, Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, IDM]
feat = 0.2088 0.1990 0.7621 0.9352 0.0031 0.0898 3.1735 0.0898 0.0080 0.9205 7.3282 0.4690 -0.0577
The result of this process is given as:
species = 'BENIGN'
And for accuracy calculation the kernel function is changed and the cross validation along with class performance is used.
The results for this process are as follows:
Accuracy of Linear kernel is: 90%
Accuracy of RBF kernel is: 80%
Accuracy of Polynomial kernel is: 73.3333%
The accuracy for different kernel emphasizes upon spread of space vector (points) in space. That is since the accuracy for linear kernel is highest it means the space vector describing the features set is linear more than radial basis function than the polynomial spread.

VIII. CONCLUSION AND FUTURE SCOPE

The world is moving more towards technology dependent era. The expertise in medical fields is as good as finding needle in haystack. Since the opinion of an expert can vary from that of novice. Hence for the lack of the expert medical practitioner in the locality should not be an obstacle for the patient of rural habitat. Hence for the benefit of all it is advisory to make the most use of the technology available to infer or conclude for treatments. The machine learning methods bring this aspect to reality, by observing the database and helping the doctor in diagnosis of diseases where lot of precision is required. And one of the machines learning technique, SVM is used in this project for classification of brain tumour to be benign or malignant. The accuracy results available range from mid 50 to 90 percent. This can be bettered by increasing the database. However the results obtained from real life images are very encouraging.

Future works and improvements
SVM, though a binary classification technique, with simple manipulation can be used for a multiple class case. This will give the medical field more space not just to classify tumour as benign or malignant. But it will give the user more options such as in case of brain images whether the image displays the properties of brain tumour or Alzheimer or is it a perfectly good image so no need to waste the time and energy in the direction that is misleading. With the proper database this method can be applied to more diseases. Example: liver diseases, skin cancer, breast cancer identification and classification etc.

IX. REFERENCES

[1]. A Kharrat, K. Gasmi, M. B. Messaoud, N. Benamrane and M. Abid, "Automated Classification of Magnetic Resonance Brain Images Using Wavelet Genetic Algorithm and Support Vector Machine", 9th IEEE Int. Conf. on Cognitive Informatics (ICCI'IO), pp. 369 - 374, 2010.

[2]. J. Huang, J. Lu, C., X. Ling, "Comparing Naive Bayes, Decision Trees, and SVM with AUC and Accuracy". Proceedings of the Third IEEE International Conference on Data Mining. IEEE Computer Society Press, pp. 553-556,2003.

[3]. Zhang, Y., L. Wu, and S. Wang, \Magnetic resonance brain image classification by an improved arti¯cial bee colony algorithm," *Progress In Electromagnetics Research*, Vol. 116, 65{79, 2011.

[4]. R. M. Haralick, K. Shanmugam and 1. Dinstein, "Textural Features of Image Classification", In Proceeding of TEEE Transactions on Systems, Man and Cybernetics, pp. 610-621, 1973.

[5]. Dr. R. J. Ramteke, KhachaneMonali Y, “Automatic Medical Image Classification and Abnormality Detection Using KNearest Neighbour”, International Journal of Advanced Computer Research (ISSN (print):2249-7277 ISSN (online): 2277-7970), Volume-2 Number-4 Issue-6. December-2012.