ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

BRAIN TUMOR CLASSIFICATION USING DATA MINING ALGORITHMS

Kalyani A. Bhawar*, Prof. Ajay S. Chhajed
*Anuradha Engineering College, Chikhli, Dist-Buldana, Maharashtra, India
Dept. of Computer Science & Engg., Chikhli, Dist-Buldana, Maharashtra, India

DOI: 10.5281/zenodo.165011

ABSTRACT

The classification of brain tumor in the magnetic resonance imaging (MRI) is very important for detecting the existence and outlines of tumors. In this paper, an algorithm about brain tumor classification is based on the metabolite values of brain MRI image is presented. Our goal is to calculate vector patterns from the metabolite values and classify the tumors automatically .Decision Trees are considered to be one of the most popular approaches for representing classifiers. Statistics, machine learning, pattern recognition, and Data Mining have dealt with the issue of growing a decision tree from available data. The purpose of this work is to present an updated survey of current methods for constructing decision tree for classifying brain tumors. The main focus is on solving the cancer classification problem using single decision tree classifiers (CART and Random algorithm).

KEYWORDS: MRI, Decision Tree, CART and Random tree Algorithm.

INTRODUCTION

Brain tumor may be a cluster of abnormal cells growing within the brain. It's going to occur in somebody at virtually any age. It's going to even amendment from one treatment session to successive however its effects might not be constant for every person. Brain tumors seem at any location, in several image intensities, will have a spread of shapes and sizes. Brain tumors may be malignant or benign. Benign brain tumors have a homogenized structure and don't contain cancer cells. They will be either monitored radiologically or surgically destroyed utterly, and that they rarely grow back. Malignant brain tumors have a heterogeneous structure and contain cancer cells. During this system, we tend to square measure getting to implement a method which may classify tumor and provides additional correct result.

In this paper the automated classification of brain magnetic resonance images by using some prior knowledge like pixel intensity and some anatomical features is proposed. Since currently there are no widely accepted methods, therefore automatic and reliable methods for tumor detection are of great need and interest. The application of decision tree systems in the classification and detection of data for MR images problems are not fully utilized yet [3]. These include the vector calculation and classification techniques especially for MR images problems with huge scale of data which consumes time and energy if done manually. Thus, fully understanding the detection, classification or vector calculation techniques is essential to the developments of decision tree systems particularly in medical-related problems. Furthermore, decision tree plays an important role in dealing with uncertainty when making decisions in medical applications [3]. Decision tree systems are systems which use Data Mining techniques in order to determine their properties by processing data samples [4, 5].

1.1. SYSTEM DESCRIPTION

There are mainly two phases in our system. Training phase and Testing phase. Overview of proposed system has been shown in Fig. 1. Initially to perform classification on MRI mages, we require image database. After gathering images we have to apply various image processing techniques in both training and testing phase. Techniques followed in these phases are, pre-processing, feature extraction, rule generation classification and Diagnosis. The pre-processing and feature extraction technique are common for both training and test phase.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Images are required to be preprocessed for feature extraction process. Extracted features are used to mine association rules for classification.

1.2 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a procedure used to produce high resolution images of the inside of the body, including the brain. MRI is a type of nuclear magnetic resonance (NMR), which is used by chemists to study the properties of molecules [3]. The technique is called magnetic resonance imaging rather than nuclear magnetic resonance imaging because of the public's negative associations with the word nuclear in the late 1970's. MRI, like other imaging techniques such as CAT and PET, is primarily a tomography imaging technique ("tomo" means slice). This means that it produces an image in a thin slice through the body. This form of imaging is, in some respects, very much like a real anatomical slice through the body. In fact, to the untrained eye, an MRI image may be indistinguishable from a slice taken from a post-mortem body. However, unlike a post-mortem slice, the MRI is formed by an MRI camera or scanner and the slices are composed of groups of picture elements or pixels displayed on a computer monitor. In MRI the subject or patient is placed inside the scanner which is made up of a moveable bed-like structure and a large hollow tube. The tube contains a thick coil of wire that generates a very intense magnetic field (between 1.5 and 4 Tesla) which is strong enough to accelerate a paper clip to near lethal velocities across the distance of a small room. In order to generate a magnetic field of this strength the coil is cooled to near absolute zero with liquid helium. This very strong magnetic field is then used to align the hydrogen nuclei of the tissue to be imaged. Also placed near the tissue to be imaged is another coil of wire called an RF coil (RF because it generates/reads radio frequency fields). The RF coil is used to both change the energy state of the hydrogen nuclei and to record the RF output of these perturbations. The latter are the raw data of an MRI which are recorded by the computer, transformed and then displayed [5].

In MRI, water molecules give off radio signals which are converted into high resolution images that look like a picture shown in figure 1.1

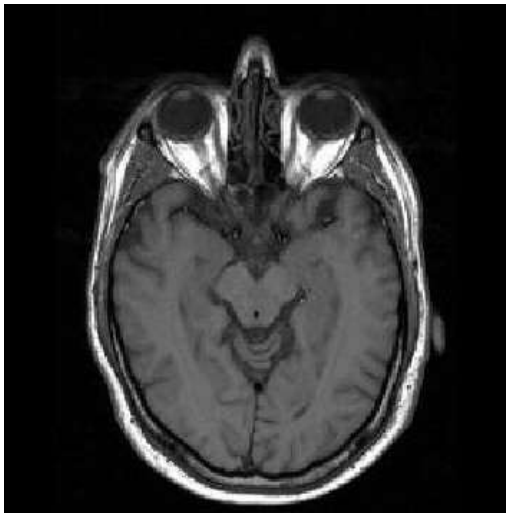

Figure 1.1: Brain MRI

PROCESS METHOD

There are 7 stages involved in the proposed model which starts from the data input to output. The first stage is the image processing system. Basically in image processing system, image acquisition and enhancement are steps that need to be done. In this project, these two steps are skipped and all the images are collected from available resources [14]. The proposed model requires extracting the metabolic values from image graph which is capable of being manipulated by the computer. The metabolic values are converted into vector form by using MATLAB. Then, the CART and RANDOM decision tree models are developed using MATLAB programming. After the CART and RANDOM decision tree model is successfully developed, the classification of the MR images starts. The proposed brain MR images classification and detection method is shown in Figure 1.2.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

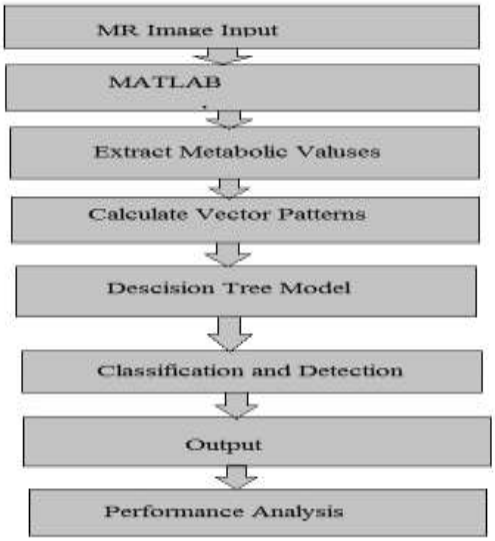

Figure 1.2: The proposed system

2.1. Metabolic Feature Extraction

The metabolite values of a patient are extracted from MR image and vector patterns are generated on the basis of metabolite values Cho/Naa, Cho, Naa, Cr2, Cr is generally simulated.

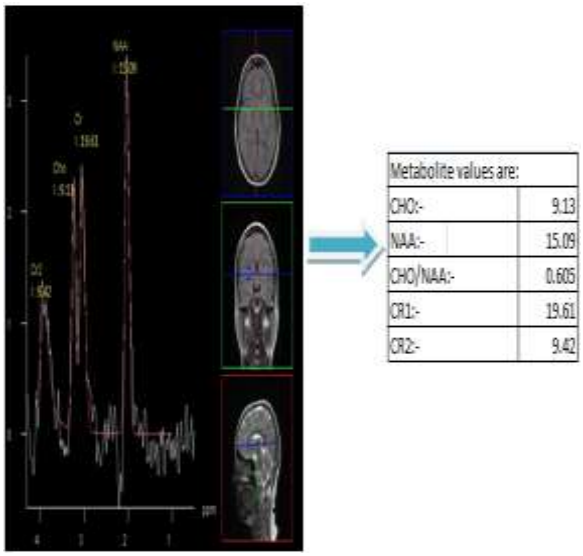

Figure 1.3: Metabolite extraction from graph image.

2.2. Vector Pattern calculation

The vector pattern calculation is based on a metabolite feature extraction. The vector pattern calculation is one of the most successful techniques that have been used in brain tumor classification and disease recognition and compression. The purpose of vector pattern calculation is to reduce the large dimensionality of the data. Brain tumor recognition systems find the disease and tumor type of a given patients test image according to their metabolite range. Vector pattern calculation and metabolites range helps to calculate the vector patterns and these are:

2.2.1. Vector patterns first bit calculation

If the metabolite cho/naa is less than 0.75 then consider the value of first bit of the vector is 1. If the metabolite cho/naa is greater than 0.75 and less than 1.25 then consider the value of first bit of the vector is 2. If the metabolite cho/naa is greater than 1.25 and less than 10.00 then consider the value of first bit of the vector is 3 else the value of first bit of the vector is 4.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

DECISION TREE

The philosophy of operation of any algorithm based on decision trees is quite simple. In fact, although sometimes containing important differences in the way to do this or that step, any algorithm of this category is based on the strategy of divide and conquers. In general, this philosophy is based on the successive division of the problem in several sub problems with a smaller number of dimensions, until a solution for each of the simpler problems can be found. Based on this principle, the classifiers based on decision trees try to find ways to divide the universe into successively more subgroups (creating nodes containing the respective tests) until each addressing only one class or until one of the classes shows a clear majority do not justifying further divisions, generating in this situation a leaf containing the class majority. Obviously, the classification is only to follow the path dictated by the successive test placed along the tree until it found a leaf containing the class to assign to the new example [10, 12, and 14]. Although the basic philosophy of all the classifiers based on decision trees is identical, there are many possibilities for its construction. Among all the key points in the selection of an algorithm to build decision trees some of them should be highlighted for their importance.

3.1. CART

CART (Classification and Regression trees) was introduced by Breiman in 1984 [22]. It builds both classifications and regressions trees. It is also based on Hunt's model of Decision tree construction and can be implemented serially. It uses gini index splitting measure in selecting the splitting attribute. Pruning is done in CART by using a portion of the training data set. CART uses both numeric and categorical attributes for building the decision tree and has in-built features that deal with missing attributes.

CART is unique from other Hunt's based algorithms as it is also used for regression analysis with the help of the regression trees. The regression analysis feature is used in forecasting a dependent variable given a set of predictor variables over a given period of time. The CART approach is an alternative to the traditional methods for prediction. In the implementation of CART, the dataset is split into the two subgroups that are the most different with respect to the outcome. This procedure is continued on each subgroup until some minimum subgroup size is reached.

EXPERIMENTAL RESULTS

The confusion matrix can be used to determine the performance of the proposed method. Here, two classification algorithms, CART and Random Decision Tree, have been implemented. This matrix describes all possible outcomes of a prediction results in table structure.

$$\text{Accuracy} = \frac{\text{Number of correct Predictions}}{\text{Total Number of Predictions}}$$

$$\text{Error rate} = \frac{\text{Number of wrong Predictions}}{\text{Total Number of Predictions}}$$

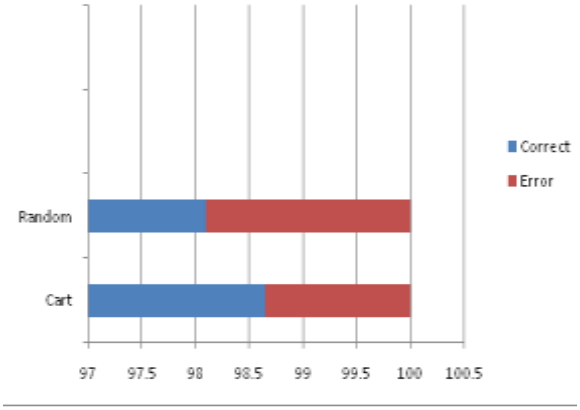

Figure1.4: Bar Chart Showing Performance of both Algorithms

CONCLUSION

"classification of brain tumor using data mining algorithm" is used to get accurate and efficient result. Using Decision tree classification technique tumor has been found as well as classified in Normal or Abnormal class. Here we used two algorithms, CART and Random Decision Tree, to compare performance. After evaluating performance we can say that the proposed algorithm has been found to be performing well compared to the existing classifiers. The accuracy of 98.5% were found in classification of brain tumor using decision tree classifier. This will produce result into normal or abnormal in efficient way. The developed brain tumor classification system is expected to provide valuable diagnosis techniques for the physicians.

REFERENCES

[1] Brain Tumor ppt. Dr. Walaa Nasr Lecturer of Medical-Surgical Nursing Department Second year 2012.
[2] K.M. Iftekharuddin, on techniques in fractal analysis and their applications in brain MRI, in: T.L. Cornelius (Ed.), Medical imaging.
[3] Jayashri Joshi1, Mrs.A.C.Phadke," Feature Extraction and Texture Classification in MRI", International Conference.
[4] Jiawei Han. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2001.
[5] Mohammed J. Zaki and Wagner Meira Jr. Fundamentals of Data Mining Algorithms. Cambridge University Press, 2010.
[6] Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques. Second edition, 2005. Morgan Kaufmann.
[7] Han J, Kamber M. Data Mining: Concepts and Techniques. Second edition, 2006. Morgan Kaufmann.
[8] Halgamuge, S. K. and L. Wang (2005). Classification and clustering for knowledge discovery. Berlin; New York, Springer.
[9] Han, J. and M. Kamber (2006). Data mining: concepts and techniques.
[10] R. Agrawal, A. Arning, T. Bollinger, M. Mehta, J. Shafer, R. Srikant: "The Quest Data Mining System", Proc. of the 2nd Int'l Conference on Knowledge Discovery in Databases and Data Mining, Portland, Oregon, August, 1996.
[11] Kearns M. and Mansour Y., On the boosting ability of top-down decision treelearning algorithms. Journal of Computer and Systems Sciences, 58(1): 109-128, 1999.
[12] Peter Mills (2011). "Efficient statistical classification of satellite measurements". International Journal of Remote Sensing.
[13] Kohavi R. and Sommerfield D., Targeting business users with decision table classifiers, in R. Agrawal, P. Stolorz & G. Piatetsky-Shapiro, eds, Proceed190
[14] Breiman, Friedman, Olshen, and Stone."Classification and Regression Trees", Wadsworth, 1984. Mezzovico, Switzerland.
[15] L. Breiman, J. Friedman., R. Olshen, C. Stone, "Classification and Regression Trees", Wadsworth, Belmont, CA.
[16] D.Steinberg., and P.L.Colla, "CART: Tree-Structured Nonparametric Data Analysis", Salford Systems: SanDiego, CA.