Disease Detection in MRS (Magnetic Resonance Spectroscopy)

Shivam Thakur

Department of Computer Engineering
Smt. Kashibai Navale College of Engineering

Sanket Patole

Department of Computer Engineering
Smt. Kashibai Navale College of Engineering

S. K. Pathan,PhD

Department of Computer Engineering
Smt. Kashibai Navale College of Engineering

ABSTRACT

Magnetic resonance spectroscopy is a technique for imaging modality used for the metabolite detection in the various parts of our body (e.g. Kidney, prostate, kidney, heart, muscle, brain) for any human being suffering from different types of disorders. It provides us the valuable information for the therapeutic monitoring of a patient as well as any diagnosis. Over the period there has been a huge amount of progress in the MRS signal processing techniques for neurometabolites quantization.

This paper presents the idea of developing a software which could be helpful to the medical experts in obtaining a classified study about diseases like brain tumor, migraine, Alzheimer's etc. with the help of magnetic resonance spectroscopy and machine learning algorithms like ID3 and Bayesian Probability. The software could easily detect the changes in the behavior of metabolites and their required functions. It will use history of patient data as training set for the application to learn on and predict the most accurate disease in the future.

General Terms

Magnetic Resonance Spectroscopy, Metabolites, MATLAB, Graphical user interface (GUI), Nuclear Magnetic Resonance Database (NMR-Db)

Keywords

Magnetic Resonance Spectroscopy, Metabolites, MATLAB, ID3 algorithm, Pandas, scikit Learn.

1. INTRODUCTION

In the MRI machine used by the medical institutions, the data is being provided in the form of .rda file format with the disturbances of the water content present in the brain. We are going to remove the disturbance of water and perform the water suppression on the FID produced through MATLAB. The removed water suppressed spectra is going to get compared by the training data of the ranges of diseases provided in a database. The coordinates of the metabolites will accurately tell us the range of the disease and better help the Experts to evaluate the chemical shifts of the metabolites of the brain.

This will help in proper detection of symptoms of diseases like tumor, Alzheimer's, Brain malformation, Dementia, Concussion, Creutzfeldt-Jakob disease. We will be comparing the metabolites like Cr, Cho, PCr, ASP, Gln, Glu, NAA, Lae. The Whole application will hide the complete Data preprocessing and finalize the result to be compared.

Magnetic resonance spectroscopy can be used to determine in a non-invasive way to find out the concentrations of the chemical substance called metabolites. The concept is based on the movement of the hydrogen atom present in the different metabolites. The spectra of MRS signal contain peaks that correspond to the different metabolites of the body.

This imaging technique is very useful while examining a patient as it reduces the amount of time and efforts required in a standard examination. This paper is about an interactive way which minimizes human labor and automates the whole process of finding a disease from a certain testing data, matching against the NMRdb. The inputs to the whole process being the .rda file which has been obtained by the clinical examination of the patients.

The proposed idea of the software would be efficiently technified to perform the preprocessing easily and methodical. This works till now has been created for the UNIX based operating system but, it only displays the water suppressed spectra.

There will be a database based on the data obtained from the training data provided in the software accompanying the idea of the proposed software. The techniques which have been used to obtain the clinical spectra involves few common terminologies which are quite useful in MRS

1.1 Field of View (FOV)

FOV is used to represent the area of different parts of the body including Brain which is under investigation. FOV is mostly expressed in standard dimensions(e.g. 21 mm, Anterior to posterior, AP; 19mm Right to Left, RL, slice thickness of 1.5 mm).

1.2 Volume of Interest (VOI)

VOI is the actual region of investigation. There are two types of voxel MRS, one being the single voxel and another being the multi-voxel. In case of a smaller voxel size, longer experiment time is required

2. METHODS

2.1 NMR DB

It is a database used for reposting the data concern- ing NMR spectra. The datasets could be easily removed and could be used for various experiments or any vari- ous need. The form in which the data is stored would be varying, ranging from the line lists stored as a graph display to raw free induction decay (FID) data.

2.1.1 Processed image.

When the FID is preprocessed then it could be easily converted into an image that would take very less memory than the compared FID. This will supply the user with a

considerably more information compared to a line list. It has less information than a raw FID, only difference being the usage of memory. This can be easily displayed in the browsers and re- quires no specialist data handling software.

2.1.2 Raw FID file.
The raw FID data obtained while per- forming the experiment would be properly stored ac- cording to the formatting interface of the instrument manufacturer. It always requires the most space but gives the maximum information. There are varieties of software programs, which will allow preprocessing the data and displaying the FID spectra.

2.2 DATA ANALYSIS
The MRS data analysis and data generation is broad- ly presented in two stages i.e. preprocessing and quantization. There are two ways in which the preprocessing steps are performed either in the time or frequency do- main. Similarly, quantization of metabolites is per- formed in either time or frequency domain.

In MRS studies, there are few conditions that impart the few imperfection in the MRS data. Some of these conditions are: - (1) The data should be collected exact- ly from the start of the FID or from the Exact center of the echo; (2) Psychotic motion; (3) Fast decaying sig- nals from immobile components; (4) overlapping arti- facts from the mobile components; (5) truncation of data before FID has decayed to noise level; and (6) Water peaks are present. Hence, there is a bigger need to preprocess the quantization of metabolites.

We have to use different techniques depending upon the route of nature of MRS data preprocessing (frequency or time domain). Some of the steps are same in both the domain, which could be resulting in similar kind of outcome. There is a direct link present in the parameters involving FID's in both the domains i.e. time and frequency. Hence the quantification of magnetic resonance signal is performed from the FID and the spectrum simultaneously.

2.3 MRS data analysis in time domain
Time domain MRS data is the actual representation of raw data, as it is also known as 'measurement do- main'. These are characterized by few parameters (resonance frequency, damping factor, individual phase and amplitude at t = 0). In the time domain signal amplitude is directly proportional to the number of molecules. The identity of the molecules is characterized by the frequency of the spectral components and the mobility of molecules is characterized by the decay constant. Appropriate model functions helps in accurate and efficient accomplishment of quantization of MRS signals. Hence it is very important step to select the appropriate model function to estimate the intrinsic parameters, which in turn gets converted into biochemical quantities (e.g. metabolite concentration).

2.4 Preprocessing
The acquired MRS data is not the ideal data due to some of the embedded imperfections from the experimental conditions. So, there is a need to perform all the preprocessing steps to eliminate the imperfections obtained. Few steps that are involved are as follows:

2.4.1 Offset correction.
The leakage in the transmitter, ref- erence frequency, which then results in non-zero mean of free induction decays, results in the DC (direct cur- rent) offset voltage.

2.4.2 Eddy current.
This imperfection occurs due to all the magnetic field gradients generated during the acquisition of the MRS signal. Time varying magnetic fields are induced which cause either frequency-dependent rephrasing, phase shifts (or signal loss) of the acquired signal.. Hence, it results in imperfect signal spectrum.

This can be recovered by dividing the phase part of the unsuppressed water signal from the water sup- pressed time domain MRS signal point-wise.

2.4.3 Noise filtering.
It is a kind of imperfection with certain standard deviation in a time domain random function. Noise is caused by either the thermal motion of the electrons in the coil or due to the receptor path. It is assumed that the noise is a Gaussian type with constant standard deviation and zero mean. Therefore noise reduction is very important in the improvement of SNR.

2.4.4 Water suppression.
The estimation of water compo- nents have been found out by applying various algorithmic approaches. The strategy, which has been found out, is by using the low pass convolution, which will suppress all the high frequency components leaving only an estimate of the water content

3. GUI(Graphical User Interface)
3.1 User Interface
The user has to enter its basic information such as Name, sex, height and weight which will then be stored in the database internally. There should be a separate text fields for all the things. The input data should be saved in the system database. The text field should be able to auto- correct the mistakes. It should be very compatible with the changes.

3.2 User Profiles
User will be defined by four different attributes-

(1) Which type of workstation they are using. (2) Their authorized user type as assigned by the system administrator. (3) Their authorized security level as assigned by the system administrator, (4) their authorized control level as assigned by the system administrator.

3.3 Security Level
The system shall support six levels of security authorization for users. The system administrator shall assign the security level when a user is registered on the system. The six levels of security enable the system to accommodate and control the actions of who have a variety of capabilities and needs. The six levels of security are (1) System administrator- highest access privilege, access to configuration tasks and all lower security level tasks. (2) Security level 1- access to all command tasks (command level 3), initialization and override commands and all lower security level tasks. (3) Security level 2- Access to macro command tasks (command level 2) and all lower security level tasks. (4) Access to super macro command tasks (Command level 1) and lower security level tasks. (5) Security level 4- no command tasks, only data entry tasks and

lower security level tasks. (6) Security level 5- review and print reports viewing system status.

3.4 Icon
The icon should be easily recognizable, after select- ing the icon the user should easily load the desired pro- gram. There should be very least ambiguity.

3.5 Communication Interface

3.5.1 Back end.
Although, being an integral part to the whole set up, the front end is responsible for automating the whole process of identifying the disease without much human intervention. Back end of the software interfaces with the NMRdB and output is given to the front end. NMRdB is a database specialized in extract- ing the chemical levels signified by the raw JPEG im- age we obtain from .rda file, which accordingly evaluates and classifies the testing data. NMRdB values are stored in MySQL database. The values of which are extracted and given to our front end JavaScript. The java script takes care of comparison (machine learning) algorithms required for finding the disease provided from the testing data.

3.5.2 Front end.
The software interfaces with the hospital staff who are supposed to fill up their data previously been saved as a testing data. The interactive GUI pro- vides assistance during the whole process until the final decision on the type of diseases is made. The user can limit the disease comparisons to a few if the doctor doesn't want to search through all the possible diseases which saves time for comparisons at back end and re- turns back with a faster output. It also provides a de- tailed view of all the chemical levels obtained from the MRS report along with the detected disease(s).

The user can opt to save the information of every tested data in corresponding user provides. The user profiles can be accessed whenever possible once saved. In case of an application crash the data is rolled back and the whole process can be started again. The application provides an additional option to report patient profiles to other hospitals through the Internet or other commu- nication protocols. There is an option to keep certain patient profile public or private

4. METHODOLOGIES
The data from NMRdb or the data directly from MATLAB is converted into .csv file using built-in py- ton/java libraries. Using the Panda library in python for data manipulations and opertaions, the data of metabolites is cleaned and prepared for classification.

Using the Scikit learn library, tools are used for statistical visualization and modeling and based on which we applied two modeling algorithms

Logicstic regression. We obtained the following results

Accuracy: 80.245% Cross-Validation Score : 80.146%

Decision trees. Or the ID3 Algorithm we achieved the following results on prediction

Accuracy: 81.930% Cross-Validation Score : 76.656%

5. CONCLUSIONS
Software was created with basic GUI to input MRS images and detect disease against past data acquired from different sources using python and the pandas library for data preparing and then applying machine learning algorithms based on it on the future on the same. Better accuracy was achieved using logistic regression keeping the cross validation score in mind. Scope for better machine learning algorithms to predict much accurate disease is expected. Cloud based system was not created given a lot of technical constraint's and a simple machine learning algorithm was used both of which are subject to further research.

6. ACKNOWLEDGEMENT

7. REFERENCES
[1] Pravat K. Mandal, In vivo proton magnetic resonance spectroscopic signal processing for the absolute quantitation of brain metabolites, European Journal Of Radiology 81 (2012) e653-e664. .

[2] Grzegorz L. Chadzynski, Uwe Klose, Chemical shift imaging without water supperession at 3T, University hospital Tuebingen,72076 tuebingen, Germany..

[3] Riddle WR, Gibbs Sj, wilcott MR, removing effects of eddy currents in proton MR spectroscopy, Med Phys 1992;19;501-509

[4] Klose U, in vivo proton Spectroscopy in presence of eddy currents. Magn reson med 1990;14:26-30.

[5] Barker PB, Soher SJ, Blackband SJ, Chatham JC, Mathew VP, Bryan RN. Quantification of proton NMR spectra of the hyuman brain using tissue water as am internal concentration reference. NMR is biomed 1993;6:89-94 .

[6] Dong Z, Dreher W, Leibfritz D towards quantitative short-echo time in vivo proton MR spectroscopy without water suppression, magn reson Med 2006;55:1441-6..

[7]Marion D, Ikura M, Bax A. Improved solvent suppression in one- and two dimensional spectra by convolution of time-domain data, J magn reson 1989;84:425-30

[8] Keeler J Understanding NMR spectroscopy . 2nd ed, U.K : John Wiley and sons; 2010. xiii, 511 p. p..

[9] Bain AD Chemical exchange in NMR. Progress in nucleur magnetic resonance spectroscopy. 2003; 43(3-4):63-103.