Application of Deep Leaning Model for Quality Control of Short-echo 7T MRSI with Various Disease Types

Huawei Liu[1], Emily Xie[1], Helene Ratiney[2], Michael Sdika[2], and Yan Li[1]

[1]Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA, United States, [2]Univ. Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, Lyon, France

Synopsis

In this project, we adapted the model with various inputs combinations, such as using different tiles, including tissue information and magnitude spectra as additional channels. These variant models were trained and tested in a comprehensive cohort of in-vivo 7T MRSI datasets of various patient types. In our test, an AUC of 0.966 was consistently achieved for the multiple type datasets.

Introduction

Magnetic resonance spectroscopic imaging (MRSI) is a noninvasive tool for assessing brain metabolites and has been applied to patients with brain tumors, multiple sclerosis, epilepsy and psychiatric disorders. The accuracy of assessing brain metabolite levels highly relies on spectra quality. Artifacts, such as field inhomogeneity, insufficient lipid suppression and spurious peaks, could interfere spectra and make the interpretation challenging. The confirmation of spectral artifacts conventionally requires manual review by MRS experts. However, due to the large number of voxels from one 3D MRSI data, this step could be time-consuming. Several recent studies have reported machine learning models are able to filter voxels with artifacts automatically[1-5]. One of these studies developed a convolutional neural network (CNN) model on 3T MRSI from patients with glioma and achieved an area under the curve (AUC) of 0.955. In this study, we evaluated the performance of this model on 7T MRSI data that were obtained from patients with different types of neurological diseases and examined features that could be included to improve its performance.

Methods

All MR data were performed using a 32-channel receive-only array with a volume transmit head coil on a GE 7T MR950 scanner. 3D MRSI data were obtained using spin-echo localization, TE/TR = 20/2000 ms, matrix size = 16-20x22x8 and nominal spatial resolution = 1cm3 and processed as described previously[6]. A total of 40 short MRSI datasets from 10 healthy controls, 10 patients with major depressive disorder, 10 patients with multiple sclerosis, and 10 patients with Parkinson's Disease. In order to balance class portions, data with subject movement were included by purpose. Voxels outside brain mask or suppressed by outer volume suppression bands were excluded. Spectra with wide spectral linewidth (FWHM>0.10ppm) and/or large Cramer-Rao lower bounds (Cr CRLB>10) were labelled as "bad". The remaining were manually examined by an expert and labeled as "bad" or "good". The final data consisted of 7,940/8,823 spectra labeled with "good/bad" and were randomly split into 8:1:1 for training, validation and testing. Additionally, segmentation of the brain was performed on T1 images with Harvard-Oxford subcortical structural atlases to get tissue ratios of cortical, subcortical, white matter, and cerebrospinal fluid. Spectra between 1.4ppm and 4.1 ppm were first standardized using sklearn.preprocessing.StandardScaler. Three types of inputs for spectra data were tested, including real spectra only, real concatenated imaginary spectra, and real concatenated magnitude spectra. Each spectrum was divided into 1, 3 or 6 tiles. Another optional input choice was to include tissue components as four individual channels. Each tile independently went through 6 convolutional and max pooling layers for feature extraction. After flattening, features from different tiles were concatenated and fed through two fully connected layers of 128 nodes then generated a scalar probability value (Figure 1). Models with different inputs and tiles were listed in Table 1. Each model was re-tested for 10 times and average accuracy (ACC) and AUC values were calculated. The pipeline was implemented in Python3 using Numpy, Pandas, Scikit-learn, and Tensorflow2 libraries.

Results and Discussions

In the study, we include datasets from healthy controls and patients with neurologic and psychiatric diseases, who had different type of metabolism. The performance of each model is illustrated in Figure 2. Our results showed adding atlas tissue information into the model did not improve the performance. There was no significant difference between different combinations of input. All models yielded similar performance at an ACC of 0.90 and AUC of 0.96-0.97, which indicated the generalizability of model for evaluating 7T MRSI data. However, several factors could limit our finding, such as the bias of one rater and two classifiers for spectral quality. Future work will include more raters for defining ground truth, evaluate inter-rater agreement, and implement Integrity Gradients[7] methods to evaluate the features importance of trained models.

Acknowledgements

References

1. Kyathanahally SP, Mocioiu V, Pedrosa de Barros N, et al. Quality of clinical brain tumor MR spectra judged by humans and machine learning tools. Magn Reson Med. 2018;79(5):2500-2510.

2. Pedrosa de Barros N, McKinley R, Wiest R, Slotboom J. Improving labeling efficiency in automatic quality control of MRSI data. Magn Reson Med. 2017;78(6):2399-2405.

3. Menze BH, Kelm BM, Weber MA, Bachert P, Hamprecht FA. Mimicking the human expert: pattern recognition for an automated assessment of data quality in MR spectroscopic images. Magn Reson Med. 2008;59(6):1457-1466.

4. Kyathanahally SP, Doring A, Kreis R. Deep learning approaches for detection and removal of ghosting artifacts in MR spectroscopy. Magn Reson Med. 2018;80(3):851-863.

5. Gurbani SS, Schreibmann E, Maudsley AA, et al. A convolutional neural network to filter artifacts in spectroscopic MRI. Magn Reson Med. 2018;80(5):1765-1775.

6. Li Y, Larson P, Chen AP, et al. Short-echo three-dimensional H-1 MR spectroscopic imaging of patients with glioma at 7 Tesla for characterization of differences in metabolite levels. J Magn Reson Imaging. 2015;41(5):1332-1341.

7. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. Proceedings of the 34th International Conference on Machine Learning - Volume 70; 2017; Sydney, NSW, Australia.
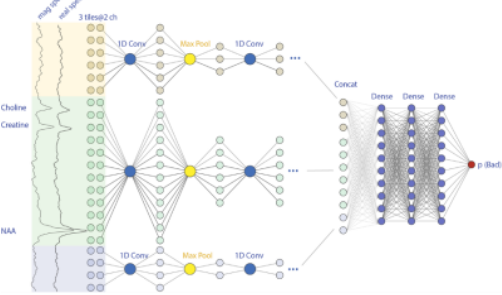
Figures

Figure 1. The convolutional neuronal network diagram for taking 3 tiles separated real and magnitude spectra as inputs. The final output value indicates the probability for bad label.

Model	Input spectra	Tiles	ACC	AUC
1	Real	1	0.906±0.008	0.969±0.005
2	Real	3	0.9±0.004	0.966±0.004
3	Real	6	0.904±0.007	0.968±0.004
4	Real & Magnitude	1	0.903±0.007	0.966±0.003
5	Real & Magnitude	3	0.899±0.007	0.964±0.003
6	Real & Magnitude	6	0.901±0.005	0.966±0.003

Figure 2. Models with different spectra inputs and tiles and the tested ACC and AUC results. No difference was found between including or not tissue ratios.