



Generating Magnetic Resonance Spectroscopy Imaging Data of Brain Tumours from Linear, Non-linear and Deep Learning Models

Nathan Olliverre¹(✉), Guang Yang^{3,4}, Gregory Slabaugh¹,
Constantino Carlos Reyes-Aldasoro², and Eduardo Alonso¹

¹ Department of Computer Science, City University of London,
London EC1V 0HB, UK

nathan.olliverre@city.ac.uk

² Department of Electrical Engineering, City University of London,
London EC1V 0HB, UK

³ Cardiovascular Biomedical Research Unit, Royal Brompton Hospital,
London SW3 6NP, UK

⁴ National Heart and Lung Institute, Imperial College London,
London SW7 2AZ, UK

Abstract. Magnetic Resonance Spectroscopy (MRS) provides valuable information to help with the identification and understanding of brain tumors, yet MRS is not a widely available medical imaging modality. Aiming to counter this issue, this research draws on the advancements in machine learning techniques in other fields for the generation of artificial data. The generated methods were tested through the evaluation of their output against that of a real-world labelled MRS brain tumor data-set. Furthermore the resultant output from the generative techniques were each used to train separate traditional classifiers which were tested on a subset of the real MRS brain tumor dataset. The results suggest that there exist methods capable of producing accurate, ground truth based MRS voxels. These findings indicate that through generative techniques, large datasets can be made available for training deep, learning models for the use in brain tumor diagnosis.

1 Introduction

Within the UK over 11,000 brain tumor cases are diagnosed each year [5]. The survival rate and period of progression have been shown to improve with treatment [22]. The process to determine treatment can be difficult and time consuming [21], this is further complicated due to limited numbers of staff available to perform these tasks [9]. These tasks also lead to the most errors in diagnosis [9]. The automation of the tasks involved in diagnosis could help to increase the accuracy and reduce the time it takes for the application of treatment to a patient.

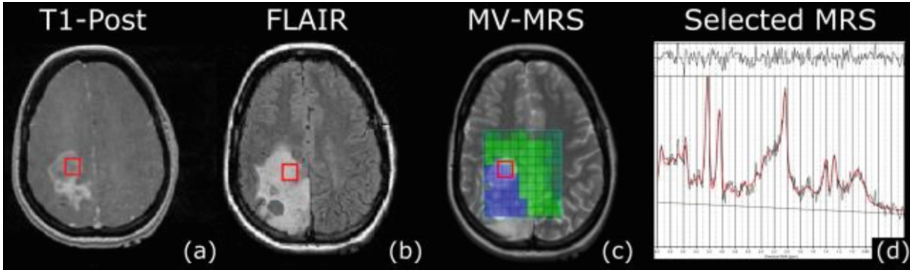


Fig. 1. Example of a T1-Post MRI (a), Fluid-Attenuated Inversion Recovery (FLAIR) MRI (b), Multi-Voxel MRS Heatmap overlaid on FLAIR (c) and the selected Single-Voxel MRS highlighted in each of the images as the red square (d) (Color figure online).

Magnetic Resonance Spectroscopy (MRS), also known as MR Spectroscopic Imaging (MRSI) or Chemical Shift Imaging (CSI), provides a non-invasive method for the diagnosis of human tissue such as lung, bone or brain matter. Similar to MR Imaging (MRI), MRS is based upon the principles of Nuclear Magnetic Resonance (NMR) [2]; however, whereas MRI uses the resultant proton signals to create detailed graphical output, MRS uses the signals to determine the quantity in parts per million (ppm) of various metabolites within cells [10] which can be seen in the example shown in Fig. 1. One area in which MRS has been shown to provide valuable insights is in the case of brain tumors [11]. In medical imaging research, MRS has proven to produce accurate results in identification of tumor grade classification [11, 20].

Current state-of-the-art computer vision models in medical imaging have begun to utilize the advancements made in machine learning, specifically in the field of deep learning [14]. These deep learning models take advantage of multi-layered networks to extract feature information from the input data. To power the ability for these deep models to extract such feature-rich information, large swaths of data are required [13]. This can be an obstacle in the medical imaging field as restrictions in data protection [7] along with non-standardized practices makes it difficult to collect the required amount of data to work with deep learning models. To be able to apply current state-of-the-art classifiers to MRS brain tumor images more data is required, hence techniques to fabricate or generate data are essential.

Advancements in computer vision and machine learning have also given the rise to accurate generative techniques for the creation of artificial data. One such generative technique is Generative Adversarial Networks (GANs) [8] which have been developed to help create more domain specific/accurate artificial data. GANs use multiple models working against each other to create accurate data, with one model (Generator) attempting to produce artificial data capable of “fooling” a model trained to determine “real” from “fake” data (Discriminator). By using approaches such as GANs the process of creating data that adheres to the domain can be achieved. Recently, variations of the original

GAN model have been developed to produce better results. The Deeply Convolutional GAN (DCGAN) [17], which takes into account the improvements that deep learning models have shown against their traditional counterparts uses a multi-convolutional layered Generator and Discriminator. Although GANs have been shown to produce accurate results in various fields [17, 19] they are known to be unstable and hard to train [3] whereas their more linear counterparts are considered to be easier to train but less expressive.

This paper applies three of the state-of-the-art methods in generating synthetic data (GAN, DCGAN and a modified MRS brain tumor classifier [16]) to the domain of MRS for review. To determine the accuracy of the artificial data created each generated dataset was used to train a Random Forest for the classification of brain tumors. The results from the trained Random Forest classifiers were bench-marked against the results of one trained on the real MRS images.

2 Materials and Methods

2.1 Materials

The MRS dataset used in this study was obtained by St. George's University London and consisted of a single-voxel MRS training and testing set. Both the test and training sets were acquired using a GE Signa Horizon 1.5T MR system with a Repetition Time (TR) and a short Echo Time (TE) of 2000 ms and 30 ms respectively. A Point-Resolved spectroscopic sequence protocol was used to acquire the training and test dataset. The World Health Organization classifies tumours into 4 grades (WHO) [6]: Grades I - IV, with GI and GII being deemed low grade tumors and GIII and GIV said to be high grade, malignant tumors. The composition of the training and test dataset was 137 samples. Of the training set 70 were classified as healthy tissue, 20 as GII (low grade), 10 as GIII and 20 GIV (high grade). The test set consisted of 9 healthy, 3 low grade and 5 high grade which were taken through random sampling of the entire dataset prior to training. Due to the similarity between GIII and GIV tumour tissue in MRS imaging, the GIII data was merged with the GIV data by labelling each as high grade. Figure 2 illustrates MRS spectra of different grades.

The positioning for every scan captured in the dataset was placed on a homogeneous, representative tumor region determined by an expert using post-Gd contrast T1w, T2w and FLAIR structural contrast images alongside the relevant histopathological information. This was to ensure accuracy within the training data and that there was a heterogeneity of MRS characteristics represented within each voxel scanned. The individual labels for the data were achieved via the diagnosis of a biopsy by a practiced physician in which the clinical, radiological and histopathological information of each patient was incorporated to the diagnosis.

The nature of MRS data is high-dimensional (of roughly 1,024 dimensions) thus a reduction with Principal Component Analysis (PCA) was applied to explore the data, see Fig. 3. Clustering with k-means results in a good separation of the classes, especially between healthy and high grade tissue (Fig. 3b).

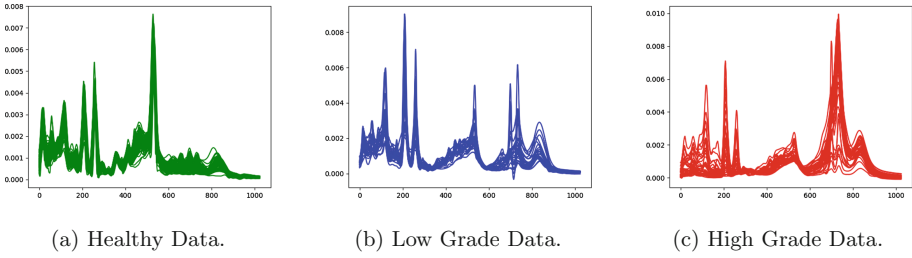


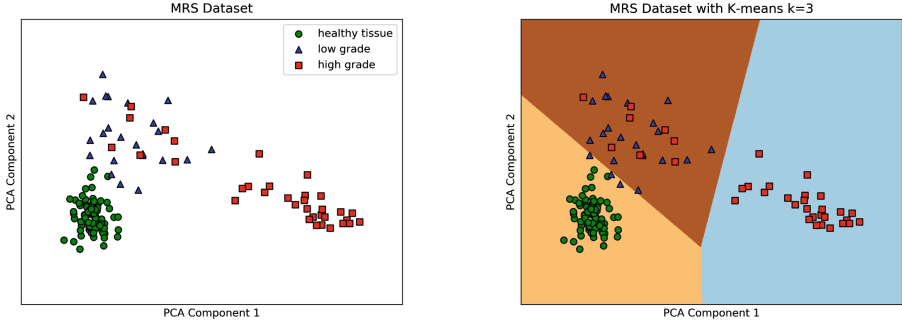
Fig. 2. Illustration of the MRS spectra grouped into a healthy (green), (b) low grade (blue) and (c) high grade (red) tumor tissue. Peaks correspond to metabolites in the region of interest. (Color figure online)

2.2 Methods

This study took a three step approach. First, three different models (GAN, DCGAN and a generative adaptation of PMM) were used to generated synthetic MRS images based off of the training -MRS- dataset. Second, these samples were then each used to train a Random Forest to be able to classify MRS images as either healthy, low or high grade tissue. Finally, the results from the Random Forest classifiers were then analyzed and the synthetic data from the generative models were compared against the mean signal of the training dataset classes.

The GAN model used in this study, based on [8], had a generator comprised of a fully connected input layer which accepted a random value vector of size 100, with a hidden layer of 1,024 nodes in size using rectified exponential linear (ReLU) [15] as the activation function with the output layer having 1,024 nodes but utilizing TanH [1] as the activation function. The discriminator consisted of a linear input layer which took in an MRS image of 1,024 values, followed by a hidden layer consisting of 1,024 nodes using LeakyReLU [1] as the activation function, the output layer was a singular sigmoidal node [18] activation. The full model was trained on each class of tissue for over 150,000 epochs using the Adam optimizer [12].

Compared to GANs, DCGANs use a deeper learning architecture which normally requires more data. To try and generate more expressive and domain accurate data, alterations were required to the training of the network to accommodate the limited data. The training process was modified to not take random batch samples but to deliberately take the full set of the data available for training at each epoch. The samples were then normalized with batch normalization. The architecture for the generator used in the study was a fully connected input layer comprised of 2048 nodes followed by four transposed one-dimensional convolutional layers using LeakyReLU as the activation function for all but the output layer which used TanH. The discriminator was the inverse with the first four layers being one-dimensional convolutions (using LeakyReLU as the activation function) followed by a single Sigmoidal output node. Similar to the GAN model the DCGAN model was trained for at least 250,000 epochs on each tissue class.



(a) PCA applied with two components.

(b) K-means clustering applied.

Fig. 3. Application of PCA to the MRS dataset with the first two components used (a) followed by k-means clustering (b).

GANs are a non-linear method for the generation of data which, theoretically, should lead to more expressive data over that of more linear methods but to do so requires larger amounts of training data. To test the benefits, a comparison to a linear based generator in the study was required. For medical imaging data, factors such as patient orientation and relationship between values matters. Therefore, testing requires a method which can generate data that adheres to the domain. The selected method for this study was the Pairwise Mixture Model.

The Pairwise Mixture Model (PMM) [16] is a model for representing different brain tissue from MRS images and based on the work by Asad et al. [4], the purpose of which was to solve the problem of the heterogeneity of tissue types found within multi-voxel MRS images. The possible types were defined as normal, low (GI and GII type brain tumour tissue) and high (GIII and GIV brain tumour tissue). Each tissue model is expressed as a mean signal and the variation around the mean, calculated from applying PCA to a labelled dataset of homogeneous MRS images each relating to a specific grade. The models could then be as defined:

$$m_i(t) = \mu_i(t) + \sum \alpha_i e_{ik}(t), \quad (1)$$

where μ_i is the mean signal, α_i and e_i respectively are the alpha weight coefficients and eigenvectors - which encode the variation around the mean signal - with K representing the number of eigenvectors determined for model m_i . To calculate the amount of each tissue type found within a certain voxel the assumption that each voxel was a weighted sum of the possible tissue types gives the following:

$$s(t) = w_n m_n(t) + w_l m_l(t) + w_h m_h(t), \quad (2)$$

which can then be viewed as an optimization problem where:

$$E = \int [x(t) - s(t)]^2 dt + \int [\sum_j s(t) - s_j(t)]^2 dt, \quad (3)$$

with j representing the available surrounding model signals of $s(t)$.

The estimated coefficients can be considered to represent the amount of each tissue type (normal, low or high) found within a voxel.

By taking the models of the various tissue types from the PMM, it is straightforward to see how by varying the value of the coefficients of the models it is possible to create data that holds to the original domain. There is a limit to the amount of possibly created data (c^3 where c represents the coefficients) but it is still enough to train a classifier model with a deep learning architecture, based on the generated data alone.

The testing of the generated data from each network used a set of simple, shallow Random Forest classifiers which were constructed and trained on the generated data from the GAN, DCGAN and the PMM generation method. A Random Forest was also trained on the training dataset alone as a control set. The Random Forest classifiers were then tested on the MRS image test set with the classification accuracy results recorded and examined. Furthermore the generated MRS images were compared to the mean signal from the training dataset for comparison to determine the adherence to the domain and possibility of expressiveness.

3 Results

Each model, the GAN, DCGAN and PMM generator, was trained on the MRS single-voxel training dataset for each class (normal, low and high). The GAN and DCGAN each had the architecture detailed in Sect. 2.2.

Table 1. Table showing the resultant classification accuracy from the set of Random Forests trained on the generated data from the GAN, PMM and the DCGAN compared to that of a Random Forest trained on only the real, labelled training dataset (GT).

Grade	PMM	GAN	DCGAN	GT
Healthy	93%	97.5%	71%	96%
Low	96%	97%	66%	95%
High	93%	93%	71%	95%

From the results shown in Table 1 the GAN produced higher accuracy in the Random Forest classifier trained on its data than with the DCGAN generated method for low grade tumor patients and for normal (healthy) tissue. The classifier trained on PMM data produced higher accuracy for low grade tumor

Table 2. Table showing the mean squared error of the linear differences between the generated signals and the mean signal of the ground truth data for each class

Grade	PMM	GAN	DCGAN
Healthy	0.002	1.021	5.148
Low	0.004	2.044	4.063
High	0.071	3.029	12.001

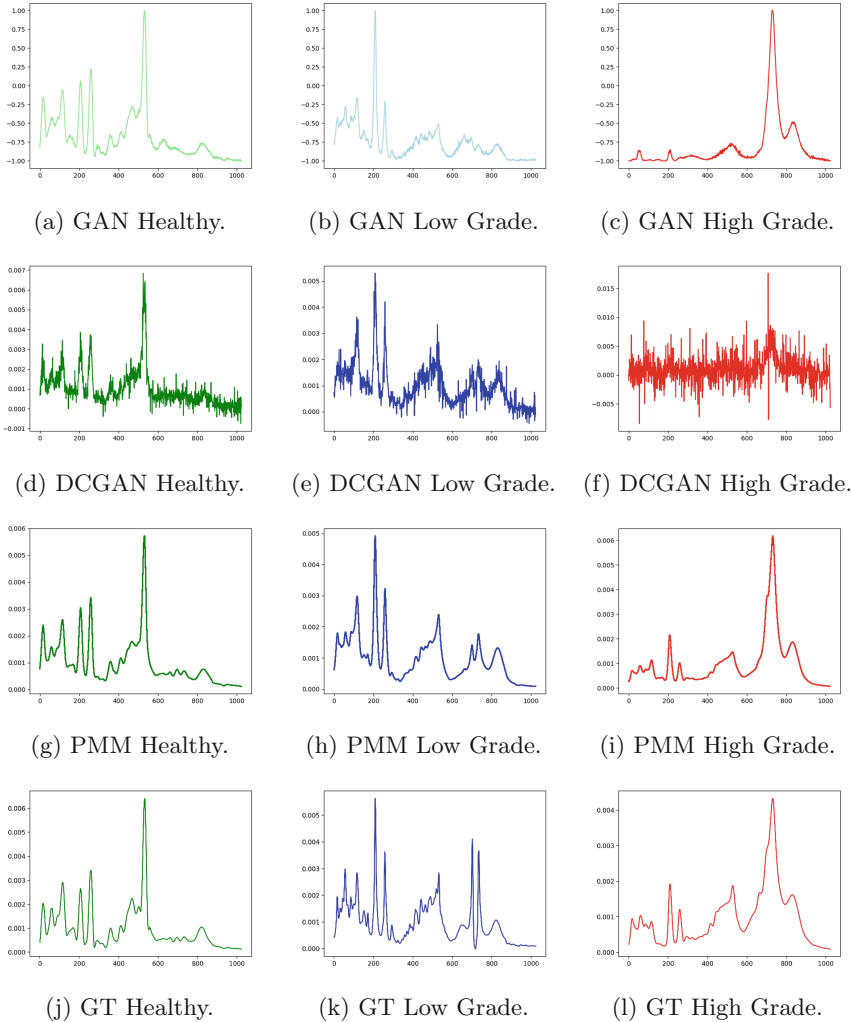


Fig. 4. Artificially generated MRS signals of healthy (green), low grade (blue) and high grade (red) tumor tissue produced from the PMM generation method (g, h, i), the GAN (a, b, c) and the DCGAN (d, e, f) compared with samples from the ground truth dataset (j, k, l). (Color figure online)

patients over that of the ground truth. The DCGAN produced the worst results over all tissue types.

The most variation within the resultant signals can be seen from both GAN methods, however, the GAN appears to cohere closer to the shape of the real MRS signals as shown in Fig. 4, this can be shown through the deltas between signals shown in Table 2.

4 Conclusions

This paper examines the possibility of generating MRS brain tumor images from limited and uneven data through one state-of-the-art generation technique, a deeper version of the same model and a modified MRS brain tumor model (GAN, DCGAN and PMM respectively). The results showed that the generated data could train a shallow Random Forest classifier to accurately determine the grades of brain tissue to the same level of one trained on real MRS data. The GAN trained model produced higher accuracy in testing and shows the expressiveness and capability in adversarial networks whereas the DCGAN model produced the lowest accuracy in classification as well as in similarity to the training data, highlighting the need for larger amounts of data when working with deep learning models. The linear model was able to produce spectra that were closer in appearance to the mean signal of the training dataset. The next stage is to examine the generated voxels by a domain expert to acknowledge their potential accuracy and expressiveness, the analysis drawn can then be used to determine what needs to be done in order to have the generated data successfully train a deep learning model for the classification of MRS images.

Acknowledgements. We would like to thank Professor Franklyn Howe at St George's, University of London, for the brain tumour MR data used in this research as well as his insights into MRS. We are also grateful for the hardware provided to us by NVIDIA that was used in this research.

References

1. Agostinelli, F., et al.: Learning activation functions to improve deep neural networks. In: [arXiv:1412.6830](https://arxiv.org/abs/1412.6830) [cs, stat], 21 December 21 2014
2. Andrew, E.R.: Nuclear magnetic resonance and the brain. *Brain Topograph.* **5**(2), 129–133 (1992)
3. Arjovsky, M., et al.: Towards principled methods for training generative adversarial networks. In: [arXiv:1701.04862](https://arxiv.org/abs/1701.04862) [cs, stat], 17 January 2017
4. Asad, M., et al.: Supervised partial volume effect unmixing for brain tumor characterization using multi-voxel MR spectroscopic imaging. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 436–439, April 2016
5. CRUK Cancer Research UK: Brain, other CNS and intracranial tumours statistics. Cancer Research UK, 14 May 2015. <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/brain-other-cns-and-intracranial-tumours>. Accessed 03 Jan 2018

6. Centre international de recherche sur le cancer et al. (eds.): Pathology and genetics of tumours of the lung, pleura, thymus and heart. IARC Press, Lyon (2004). OCLC: 492146534
7. Cios, K.J., et al.: Uniqueness of medical data mining. *Artif. Intell. Med.* **26**(1), 1–24 (2002). Medical Data Mining and Knowledge Discovery
8. Goodfellow, I., et al.: Generative adversarial nets. In: Ghahramani, Z., et al. (eds.) *Advances in Neural Information Processing Systems 27*, pp. 2672–2680. Curran Associates Inc. (2014)
9. Grant, R.: Overview: brain tumour diagnosis and management/Royal College of Physicians guidelines. *J. Neurol. Neurosurg. Psychiat.* **75**(Suppl. 2), ii18–ii23 (2004)
10. Gujar, S.K., et al.: Magnetic resonance spectroscopy. *J. Neuro-Ophthalmol.* **25**(3), 217 (2015)
11. Howe, F.A., et al.: 1H MR spectroscopy of brain tumours and masses. *NMR Biomed.* **16**(3), 123–131 (2003)
12. Kingma, D.P., et al.: Adam: a method for stochastic optimization. In: [arXiv:1412.6980](https://arxiv.org/abs/1412.6980), 22 December 2014
13. LeCun, Y.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
14. Litjens, G., et al.: A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017)
15. Nair, V., et al.: Rectified linear units improve restricted Boltzmann machines. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML 2010*, pp. 807–814. Omnipress, Madison (2010)
16. Olliverre, N., et al.: Pairwise mixture model for unmixing partial volume effect in multi-voxel MR spectroscopy of brain tumour patients. In: *Medical Imaging 2017: Computer-Aided Diagnosis*, 3 March 2017, vol. 10134. International Society for Optics and Photonics (2017). 101341R
17. Radford, A., et al.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: [arXiv:1511.06434](https://arxiv.org/abs/1511.06434) [cs], 19 November 2015
18. Rumelhart, D.E.: Learning representations by back-propagating errors. *Nature* **323**(6088), 533–536 (1986)
19. Salimans, T., et al.: Improved techniques for training GANs. In: [arXiv:1606.03498](https://arxiv.org/abs/1606.03498) [cs], 10 June 2016
20. Sibtain, N.A.: The clinical value of proton magnetic resonance spectroscopy in adult brain tumours. *Clin. Radiol.* **62**(2), 109–119 (2007)
21. Yang, G.: Discrete wavelet transform-based whole-spectral and subspectral analysis for improved brain tumor clustering using single Voxel MR spectroscopy. *IEEE Trans. Biomed. Eng.* **62**(12), 2860–2866 (2015)
22. Zacharaki, E.I.: Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn. Reson. Med.* **62**(6), 1609–1618 (2009)