

# 3D Texture Analysis of Heterogeneous MRI Data for the Characterisation of Childhood Brain Tumours

Ahmed E Fetit<sup>1,2</sup>, Jan Novak<sup>2,3</sup>, Daniel Rodriguez<sup>4</sup>, Dorothee P Auer<sup>4,5</sup>, Chris A Clark<sup>6</sup>, Richard G Grundy<sup>4,5</sup>, Tim Jaspán<sup>5</sup>, Andrew C Peet<sup>2,3</sup>, and Theodoros N Arvanitis<sup>1,2</sup>

<sup>1</sup>Institute of Digital Healthcare, WMG, University of Warwick, Coventry, United Kingdom, <sup>2</sup>Birmingham Children's Hospital, Birmingham, West Midlands, United Kingdom, <sup>3</sup>University of Birmingham, Birmingham, West Midlands, United Kingdom, <sup>4</sup>University of Nottingham, Nottingham, United Kingdom, <sup>5</sup>University Hospital Nottingham, Nottingham, United Kingdom, <sup>6</sup>University College London, London, United Kingdom

## Target Audience

This work is targeted at individuals interested in developing pattern recognition models that can quantify Magnetic Resonance Imaging (MRI) information below human visual perception, to assist with the diagnosis of childhood brain tumours.

## Purpose

Whilst conventional MRI provides images of superb clarity, the subjective nature in which assessment is performed can lead to errors due to difficulty in visually inspecting textural patterns on MRI scans. There is an increasing interest in developing quantitative MR image analysis tools that can assist with the diagnosis of childhood brain tumours. Promising results have previously been obtained from a preliminary study conducted using traditional (2D) texture analysis of a small cohort of T2-weighted MR images<sup>1</sup>. In this work, our investigation is extended by carrying out 3D (volumetric) texture analysis on a multimodal, heterogeneous MR data set. 3D techniques are applied over several adjacent slices along the X, Y and Z axes. It is hypothesised that extracting 3D texture features across multiple slices would yield more information about the tumours than applying 2D techniques on only one slice. The volumetric texture information can then be used to build a quantitative model that can differentiate between Medulloblastoma (MB), Pilocytic Astrocytoma (PA) and Ependymoma (EP), which are the most common types of brain tumours in children.

## Materials and Methods

- **Dataset:** In this study we used anonymised T1 and T2 weighted MR images of 136 children diagnosed with Medulloblastoma (74), Pilocytic Astrocytoma (44) and Ependymoma (18). The dataset is 'heterogeneous' in the sense that the number of cases is imbalanced and the scans were acquired from 3 different hospitals across the UK. The following work was conducted on the axial slices of the pre-contrast T1 and T2 images.
- **Pre-processing:** Snake GVF segmentation was carried out on MATLAB in order to extract regions of interest (ROIs) in which the tumour was present. Normalisation ( $\mu/\sigma$ ) was carried out<sup>2</sup> on the ROIs in order to mitigate different acquisition protocols across the hospitals.
- **Texture Analysis:** MaZda software package<sup>3</sup> was then used to separately carry out 2D and 3D texture analysis. The analysis was based on the following techniques: histogram, absolute gradient, grey-level co-occurrence matrix, grey-level run-length matrix, wavelets and autoregressive model. The principles behind these techniques are well explained by Castellano et al.<sup>4</sup>
- **Minority Over-Sampling:** In order to address the imbalanced nature of our dataset, Synthetic Minority Over-Sampling Technique (SMOTE) was applied to the extracted 3D features<sup>5</sup>. SMOTE was used to create 39 synthetic Ependymoma samples by operating in feature space.
- **Feature Analysis:** The entropy-MDL discretisation algorithm<sup>6</sup> was then used to partition our (continuous) features to a discrete number of intervals and for feature selection. The three discretised feature vectors (2D, 3D and 3D\_SMOTE) were separately analysed using four supervised machine-learning (ML) algorithms: Naive Bayes, k-Nearest Neighbour, Classification Tree and Support Vector Machine (SVM). Classification performance was evaluated using the Leave-One-Out validation approach

## Results

SVM has yielded the highest classification accuracy results for all three feature vector groups, achieving 69%, 72% and 81% with 2D, 3D and 3D\_SMOTE respectively. The dataset with synthetic samples (3D\_SMOTE) has achieved exceptionally high sensitivity results when classifying Ependymomas. Table 1 shows a summary of the results achieved by SVM.

## Conclusions

The results support the use of texture analysis as an automated, quantitative technique to assist with the diagnosis of paediatric brain tumours. Training the classifiers with 3D textural features has improved the classification accuracy as hypothesised, achieving 72% with SVM (3% improvement from the results obtained with 2D features). Our findings also show that the use of synthetic Ependymoma samples has helped the classifier build a more 'realistic' model of the features, achieving an extraordinary increase in Ependymoma sensitivity (92% vs. 0%). Moreover, the use of synthetic samples has led to an increase of the overall classification accuracy (82% vs. 72%).

Feature Set	Classification Accuracy	Medulloblastoma		P. Astrocytoma		Ependymoma	
		Sens	Spec	Sens	Spec	Sens	Spec
2D	69	67	80	89	65	0	99
3D	72	69	80	92	70	0	99
3D SMOTE	82	57	96	82	88	92	87

Table 1 Results obtained with SVM on the three datasets (Leave-One-Out).

## References

1. A. E. Fetit, J. Novak et al MRI texture analysis in paediatric oncology: a preliminary study, *Stud in Health Tech & Info* (2013) 130: 169-171
2. G. Collewet et al, Influence of MRI acquisition protocols and image intensity normalisation methods on texture classification *MRI* (2004) 22: 81-91
3. P. Szczypinski et al, MaZda: a software package for image texture analysis, *Computer Methods & Programs in Biomedicine* (2009), 94(1): 66-76
4. G. Castellano et al, Texture analysis of medical images, *Clinical Radiology*, Volume 59, Issue 12, Pages 1061-1069
5. N.V. Chawla et al, SMOTE: Synthetic Minority Over-Sampling Technique, *Artificial Intelligence Research* (2002) 16:321-357
6. L. Jinyan et al. Mean-entropy discretized features are effective for classifying high-dimensional biomedical data. *3rd ACM SIGKDD Workshop on Data Mining* (2003).