

# Wavelet oversampling for imbalance childhood brain tumor classification

Zhao, Teddy; Grist, James; Rose, Heather; Sun, Yu; Peet, Andrew

*License:*

None: All rights reserved

*Document Version*

Early version, also known as pre-print

*Citation for published version (Harvard):*

Zhao, T, Grist, J, Rose, H, Sun, Y & Peet, A 2021, Wavelet oversampling for imbalance childhood brain tumor classification. in *Proceedings of the 2021 ISMRM & SMRT annual meeting & exhibition.*, 0937, Proceedings of the International Society for Magnetic Resonance in Medicine, Scientific Meeting and Exhibition, International Society for Magnetic Resonance in Medicine, 2021 ISMRM & SMRT Annual Meeting & Exhibition, 15/05/21. <<https://cds.ismrm.org/protected/21MProceedings/PDFfiles/0937.html>>

[Link to publication on Research at Birmingham portal](#)

**General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

**Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# Wavelet Oversampling for Imbalance Childhood Brain Tumour Classification

Dadi Zhao<sup>1,2</sup>, James T. Grist<sup>1,2</sup>, Heather E.L. Rose<sup>1,2</sup>, Yu Sun<sup>1,2</sup>, and Andrew C. Peet<sup>1,2</sup>  
<sup>1</sup>Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, United Kingdom, <sup>2</sup>Department of Oncology, Birmingham Children's Hospital, Birmingham, United Kingdom

## Synopsis

**Classifying imbalance childhood brain tumours through <sup>1</sup>H-MRS metabolite profiles remains a challenging problem. We presented an alternative oversampling method, wavelet oversampling (WvOS). Different from the classic Synthetic Minority Oversampling TEchnique that oversamples the metabolite profiles, WvOS used the wavelet processed <sup>1</sup>H-MRS as the oversampled <sup>1</sup>H-MRS, followed by quantification and classification. As the result, WvOS can provide dramatically better classification performance than non-oversampled or classic oversampled metabolite profiles. An optimal balanced classification accuracy is achieved as 96% and 72% from 84% and 52% for the 1.5T and 3T cohorts of childhood brain tumours, respectively.**

## Introduction

Combined with machine learning, metabolite profiles provide support for childhood brain tumour diagnosis<sup>1</sup>. The classification performance of brain tumours showed promising for common types like medulloblastomas and pilocytic astrocytomas but limited for rare and diverse types like ependymomas. Such imbalance multi-class classification problems could be addressed by oversampling for the minority, which was previously reported by using Synthetic Minority Oversampling TEchnique (SMOTE)<sup>2-3</sup>. In this abstract, we presented an alternative oversampling method, Wavelet OverSampling (WvOS), which is based on the wavelet decomposition and reconstruction. Compared with SMOTE, WvOS can provide stably better classification performance and could guide further advances of brain tumour classification through deep learning.

## Theory

The WvOS (Figure 1) is an extension of the Wavelet Noise Suppression (WvNS)<sup>4</sup> for proton magnetic resonance spectroscopy. Both WvNS and WvOS assume that some level of noise is presents in the <sup>1</sup>H-MRS signals. WvNS generates temporary <sup>1</sup>H-MRS signals from the original <sup>1</sup>H-MRS signals by using a series of wavelet bases. WvOS selects high-quality <sup>1</sup>H-MRS signals from these temporary <sup>1</sup>H-MRS signals as the results of oversampling. Specifically, the temporary <sup>1</sup>H-MRS signals with the best and the second best signal quality is selected as the noise-suppressed <sup>1</sup>H-MRS signal and the oversampled <sup>1</sup>H-MRS signal, respectively. If higher oversampling rate is preferred, the temporary <sup>1</sup>H-MRS signal with the third best signal quality and the following are used.

## Methods

Pre-operative patients presenting with a brain tumour like ependymoma, medulloblastoma or pilocytic astrocytoma were recruited in this retrospective study. Clinical <sup>1</sup>H-MRS scanning was performed on a 1.5T or 3T scanner in Birmingham Children's Hospital (TE 30-46ms, TR 1500-2000ms, voxel size 20-80mm<sup>3</sup>). The raw <sup>1</sup>H-MRS were quantified by using TARQUIN (version 4.3.11) with a <sup>1</sup>H-brain full basis and screened by two experienced spectroscopists according to the quality of the spectrum. Overall signal-to-noise ratio (SNR) was used to define the quality of <sup>1</sup>H-MRS signals. The <sup>1</sup>H-MRS analysis and brain tumour classification were conducted separately for the 1.5T and 3T cohorts. Metabolite profiles were used as the features for brain tumour classification. Noise suppression for <sup>1</sup>H-MRS was conducted through WvNS with one level of decomposition in the frequency domain<sup>4</sup>. Metabolites were ranked and selected based on their multi-class area under the curve, and the number of metabolites used for classification was no more than the number of smallest tumour group. Oversampling for the minority was conducted through WvOS and compared to SMOTE. Classification experiments were conducted through linear discriminant analysis and support vector machine with a linear kernel. The classification accuracy was determined through leave-one-out and ten-fold cross validation and compared between non-oversampled and oversampled post-noise suppression <sup>1</sup>H-MRS through a Wilcoxon signed-rank test.

## Results

Totally 82 patients were enrolled to the 1.5T cohort (ependymomas, *N*=12; medulloblastomas, *N*=31; pilocytic astrocytomas, *N*=39), and 20 patients were enrolled to the 3T cohort (ependymomas, *N*=2; medulloblastomas, *N*=8; pilocytic astrocytomas, *N*=10). The classification performance of the oversampled metabolite profiles showed significantly better than the non-oversampled metabolite profiles (*P*<.05, Figure 2-3), among which WvOS provides significantly better classification performance than non-oversampled or SMOTE-

## Figures

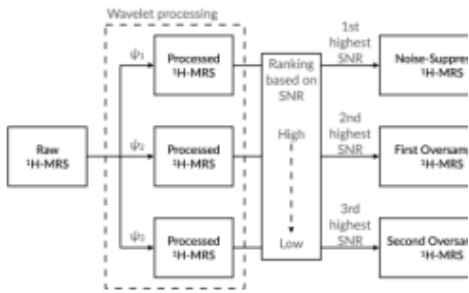
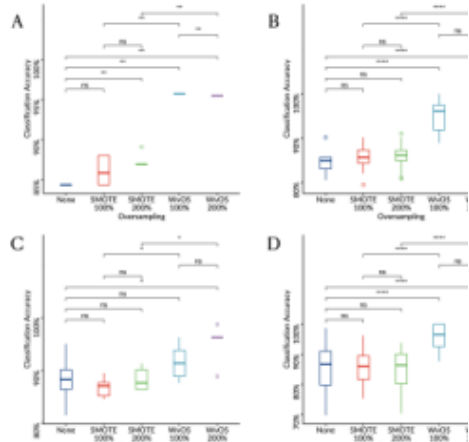
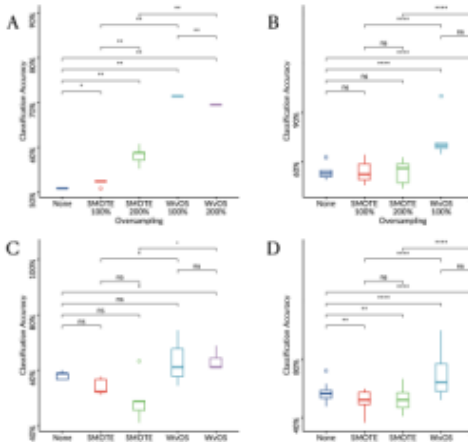


Illustration showing the procedure of generating oversampled proton magnetic resonance spectroscopy (<sup>1</sup>H-MRS) from the raw <sup>1</sup>H-MRS by using Wavelet OverSampling (WvOS). Abbreviations: SNR, signal-to-noise ratios;  $\psi$ , wavelet basis.



Boxplots showing the balanced classification accuracy for the 1.5T cohort of childhood brain tumours derived through linear discriminant analysis (A, C) or support vector machine (B, D) with leave-one-out (A-B) or six-fold (C-D) cross validation.



Boxplots showing the balanced classification accuracy for the 3T cohort of childhood brain tumours derived through linear discriminant analysis (A, C) or

based oversampled metabolite profiles ( $P<.05$ ). For example, according to linear discriminant analysis and leave-one-out cross validation, WvOS showed the optimal balanced classification accuracy as 96% for the 1.5T cohort (Figure 2) and 72% for the 3T cohort (Figure 3), which is significantly improved from 84% for the 1.5T cohort and 52% for the 3T cohort before oversampling, and 87% for the 1.5T cohort and 58% for the 3T cohort through SMOTE, respectively ( $P<.05$ ).

## Discussion

Childhood brain tumour classification through metabolite profiles in clinical diagnosis has met challenges including the limited signal quality, the performance of fitting, the selection of metabolites and the imbalance group size, which has led to limited classification accuracy of childhood brain tumours. Recent progress has shown improved classification accuracy through novel feature selection and noise suppression<sup>5</sup>, while the issue of limited group size for the rare tumour types like ependymomas remains to be investigated.

SMOTE is a useful tool to oversample the minority to reduce the negative affect of imbalance group size<sup>3</sup>. However, SMOTE can be used for metabolite profiles instead of the <sup>1</sup>H-MRS, which makes it vulnerable to the limited fitting performance and could generate unpredictable and inaccurate results. In contrast, the WvOS used in-depth processing on the raw <sup>1</sup>H-MRS and guaranteed the reliability of oversampled results.

Due to the presence of noise, the ground-truth of metabolite profiles remain undetectable, thus the data reliability is similar between the noise-suppressed and oversampled <sup>1</sup>H-MRS. Due to the difficulty of clinical data collection, deep learning was rarely reported in classifying childhood brain tumours through <sup>1</sup>H-MRS. Benefited from the rich variation of wavelet profiles, it is possible to generate a large-scale <sup>1</sup>H-MRS dataset from a small clinical cohort. Therefore, the current method has the potential to guide the application of deep learning in classifying childhood brain tumours through <sup>1</sup>H-MRS.

## Conclusion

Wavelet oversampling provides an alternative solution with better performance for imbalance classification of childhood brain tumour classification.

## Acknowledgements

We would like to acknowledge funding from the Cancer Research UK and EPSRC Cancer Imaging Programme at the Children's Cancer and Leukaemia Group (CCLG) in association with the MRC and Department of Health (England) (C7809/A10342), the Cancer Research UK and NIHR Experimental Cancer Medicine Centre Paediatric Network (C8232/A25261), Health Data Research UK (HDR UK) and the Children's Research Fund.

## References

1. Davies NP, Wilson M, Harris LM et al. Identification and characterisation of childhood cerebellar tumours by in vivo proton MRS. *NMR in Biomedicine*. 2008;21:908-918.
2. Zarinabad N, Wilson M, Gill SK, Manias KA, Davies NP, Peet AC. Multiclass imbalance learning: Improving classification of pediatric brain tumors from magnetic resonance spectroscopy. *Magnetic Resonance in Medicine*. 2017;77:2114-2124.
3. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. 2002;16:321-357.
4. Zhao D, Grist JT, Sun Y, Peet AC. Impact of wavelets and apodisation in magnetic resonance spectroscopy quality for paediatric brain tumours. *Proceedings of the Annual Meeting of ISMRM, Montréal, Canada*. 27: 4243.
5. Zhao D, Grist JT, Sun Y, Sawlani V, Peet AC. Optimised paediatric brain tumor diagnosis by using in vivo magnetic resonance spectroscopy and machine learning. *Proceedings of the Annual Meeting of ISMRM, Sydney, Australia* 28: 1394.

support vector machine (B, D) with leave-one-out (A-B) or six-fold (C-D) cross validation.