



HAL
open science

Multiple Excitation Wavelengths PpIX Fluorescence Spectroscopy for Gliomas Delineation Assisted by Machine Learning

Hermine Quardon, Arthur Gautheron, Antoine Uzel, David Meyronet, Bruno Montcel, Thiébaud Picart, Cédric Ray

► To cite this version:

Hermine Quardon, Arthur Gautheron, Antoine Uzel, David Meyronet, Bruno Montcel, et al.. Multiple Excitation Wavelengths PpIX Fluorescence Spectroscopy for Gliomas Delineation Assisted by Machine Learning. ECBO - European Conferences on Biomedical Optics, OPTICA; SPIE, Jun 2025, Munich (Allemagne), France. pp.101, <10.1364/ECBO.2025.W2A.3>. <hal-05069013>

HAL Id: hal-05069013

<https://hal.science/hal-05069013v1>

Submitted on 15 May 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Multiple Excitation Wavelengths PpIX Fluorescence Spectroscopy for Gliomas Delineation Assisted by Machine Learning

Hermine Quardon^{1,*}, Arthur Gautheron^{1,2}, Antoine Uzel¹, David Meyronet^{3,4,5}, Bruno Montcel¹,
Thiébaud Picart^{4,5,6}, and Cédric Ray^{1,**}

1 Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F69100 Lyon, France

2 CPE, Lyon, France

3 Department of Neuropathology, Groupement Hospitalier Est, Hospices Civils de Lyon, 69677 Bron, France

4 Faculty of Medicine, University Claude Bernard Lyon 1, 69100 Villeurbanne, France

5 Cancer Initiation and Tumoral Cell Identity Department, Cancer Research Centre of Lyon (CRCL) INSERM 1052, CNRS 5286, 69008 Lyon, France

6 Department of Neurosurgery, Hôpital Neurologique Pierre Wertheimer, Hospices Civils de Lyon, 69677 Bron, France

*: hermine.quardon@creatis.insa-lyon.fr

** : cedric.ray@univ-lyon1.fr

Abstract: A method for preprocessing and classifying fluorescence measurements to distinguish tumorous from healthy tissue is proposed, as 5-ALA fluorescence-guided surgery, effective for high-grade gliomas, is less useful for low-grade gliomas without visible fluorescence. © 2025 The Authors

1. Introduction

Gliomas are invasive tumors with unclear boundaries, making it challenging to distinguish them from the surrounding healthy tissue. The World Health Organization (WHO) classifies gliomas into four grades [1]; however, most research typically groups them into two categories: High-Grade Gliomas (HGG) and Low-Grade Gliomas (LGG). Enhancing the extent of tumor resection is crucial in reducing the risk of recurrence and improving both quality of life and survival rates.

Fluorescence of protoporphyrin IX (PpIX) induced by 5-aminolevulonic acid (5-ALA) is currently used as PpIX emits a detectable pink fluorescence in tumoral tissue when illuminated by a blue light source from the surgical microscope. However, its sensitivity remains limited [2]. Several methods using 5-ALA-induced PpIX fluorescence spectroscopy have been proposed to address these sensitivity challenges. The general approach involves modeling a fluorescence baseline, which includes all components not related to 5-ALA-induced PpIX, and is assumed to contain a finite number of fluorophores with known fluorescence spectral profiles. Studies demonstrate that the complexity of PpIX fluorescence emission spectra in tissue is closely associated with the pathological condition [3,4].

This study aims to develop a machine learning-based tool to assist neurosurgeons in real-time by accurately selecting biopsy sites. The machine learning helps differentiate tumorous from healthy tissue, improving surgical precision and outcomes. To do so, we use PpIX fluorescence spectra, which are acquired with two different excitation wavelengths. These wavelengths range from 375 nm to 440 nm, which are within the PpIX absorption spectrum [5]. Assuming that the baseline is patient-dependent and affects the PpIX intensity, a normalization of each spectrum by this quantity can lead to a better distinction between tumor margin and healthy samples. The results indicate that this preprocessing of the data leads to a high-quality classification for both HGG and LGG data.

2. Materials and Methods

2.1. Data Sources

This study used a custom acquisition device, described in [3] as System B. The dataset was collected as part of a clinical study conducted at the Neurological Center of Pierre Wertheimer Hospital in Bron, France. All experiments were in accordance and approved by the French Agency for Health (ANSM) and the local ethics committee of Lyon University Hospital (France). All participating patients provided written informed consent.

Patients were given an oral dose of 5-ALA to induce PpIX fluorescence. For each patient, the surgeon exposed the tissue and selected the area to resect. Then, two samples were extracted in several sites: the first sample was sent to histopathological analysis and the second one was used for this study. Each acquisition lasted 200 ms, with the laser turned on, followed by the same duration with the laser turned off to suppress the ambient light coming from the operating room. Ten measurements were taken for each sample and then averaged. These fluorescence

measurements were performed in order to get representative spectra from areas with different densities of tumor cells. In the first step, the spectra were acquired with the optical system described above. Two fluorescence spectra corresponding to two excitation lasers were acquired. The pathological status of the corresponding tissue was determined by the histopathological analysis, giving each sample a label. In total, 169 measurements were kept in this analysis, including 98 samples of High-Grade Gliomas (HGG), 59 samples of Low-Grade Gliomas (LGG) and 12 samples of Healthy tissue.

2.2. Data Preprocessing

The data preprocessing pipeline developed is illustrated in Fig. 1 and is based on the most commonly used steps of spectral data processing in neuro-oncology according to [6].

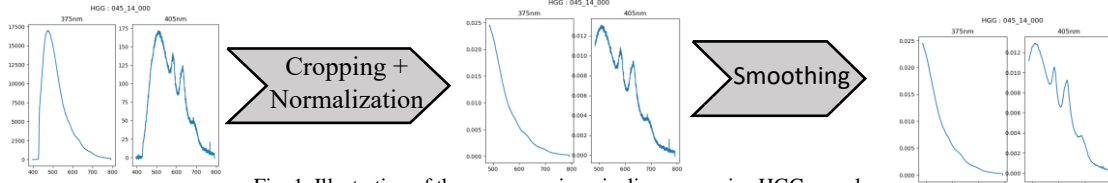


Fig. 1. Illustration of the preprocessing pipeline on a noisy HGG sample.

The spectra were first cropped and then normalized by the contribution of endogenous fluorophores in the spectra. The wavelength of the crop and the wavelength to which the signal was considered autofluorescence were empirically determined and respectively set to 485nm and 530nm. The normalization was followed by a third order polynomial Savitsky-Golay filtering with a 20nm window length, applied on every normalized spectrum to smooth it. Finally, the preprocessed spectra from the two excitations were concatenated to group all the information of the sample.

2.3. Feature Extraction

The feature space created by the concatenation of the two spectra was too large to be efficiently applied to classification methods. Dimension reduction techniques were used to reduce the redundancy in the concatenated spectrum. Two data-based feature projection techniques were tested, the Principal Component Analysis (PCA) [7] and the Linear Discriminant Analysis (LDA) [8]. While the PCA is an unsupervised technique, the LDA is supervised and also provides a classification.

PCA was mainly chosen for its ability to decompose the data according to their physical properties. The number of principal components is determined by the minimum number of components to reach a 95% cumulated variance. Finally, LDA was chosen to test the influence of the labels in the feature extraction.

2.4. Data Augmentation

As the size of the dataset is relatively small and classes are unbalanced, a data augmentation technique was added after/before reducing the dimensions in PCA/LDA pipelines respectively, the Synthetic Minority Over-sampling Technique (SMOTE) [9]. The augmentation was performed mainly to enable a fine-tuning of the machine learning classification models.

This data augmentation was also performed to increase the trustworthiness of the validation method and the evaluation of the model, as too few samples were available to obtain a representative metric.

2.5. Machine Learning Models

A binary classification from the reduced feature space was then performed. Unlike [4], a distinction between HGG and LGG was made, as the PpIX fluorescence highly depends on the grade of the tumor [5]. Then two separate binary classification models were tuned independently for high- and low-grade gliomas.

Two supervised models were tested in this study. As LDA was used as a dimension reduction technique, it was consistent to test its abilities also as a classifier, especially since it does not require numerous hyperparameters to adjust. The second model presented was a K Nearest Neighbors (KNN) classifier which allowed to tune some hyperparameters such as the number of neighbors or the weight of those neighbors in the classification decision.

2.6. Validation and Model Evaluation

To avoid an overfitting of the model, a KFold cross-validation was performed. The data available were split in 10 smaller data sets. As this study also aimed to assess the relevance of the data augmentation in the unbalanced classes issue, all the metrics used took into account the imbalance between healthy and tumor samples. The models were evaluated through balanced accuracy and weighted F1-score.

3. Results and Discussion

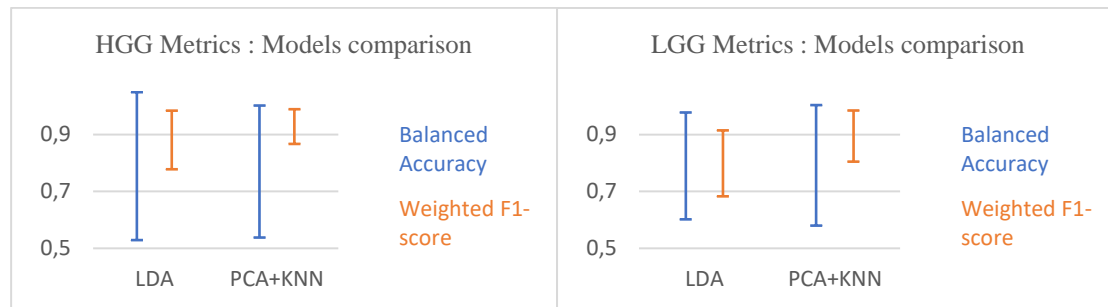


Fig. 2. Metrics of the two tested classification models for high (a) and low (b) grades: LDA and PCA+KNN.

It appears in Fig. 2 that both supervised classification models can reach the same high quality of results, even for LGG data. A comparison between results obtained with the supervised methods and an unsupervised clustering method could be performed.

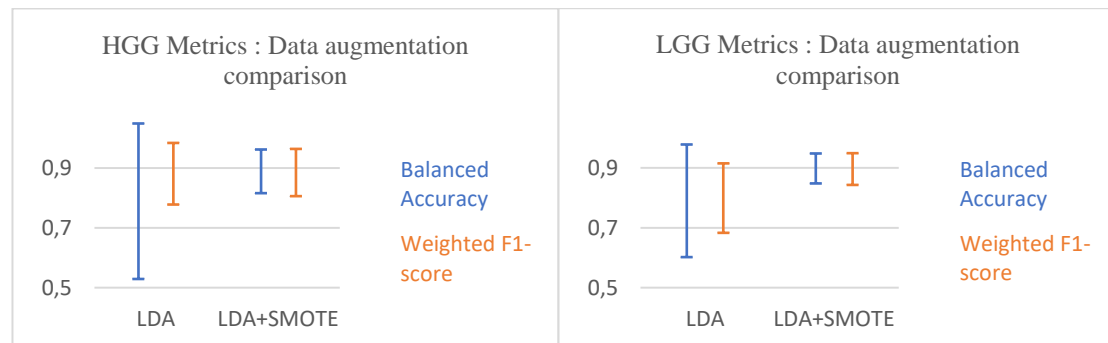


Fig. 3. Metrics of the LDA model without and with augmentation for high (a) and low (b) grades.

As shown in Fig. 3, the data augmentation significantly increased the robustness of the models which means that incorporating new acquisitions in the data sets will improve the classification and a better delineation of glioma margins. As suggested in [6], an attenuation correction will be added in the preprocessing pipeline, in order to reduce the blood absorption in the spectra.

Acknowledgements:

This work was supported by the European Union's Horizon Europe research and innovation program under grant agreement No 101071040. This work was also supported by the Graduate School Medical Device Engineering of the Université Claude Bernard Lyon 1 within the program "Graduate +" (21-SFRI-0001) operated by the French National Research Agency (ANR), by LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007), operated by the French National Research Agency (ANR) and by France Life Imaging (ANR-11-INBS-0006). This material is based on work done on the PILoT facility (PILoT, INSA LYON, Bât. L. de Vinci, 7 Av. Jean Capelle 69621 Villeurbanne).

References:

- [1] A. Sejda, W. Grajkowska, J. Trubicka, E. Szutowicz, T. K. Wojdacz, W. Kloc et al. "WHO CNS5 2021 classification of gliomas: a practical review and road signs for diagnosing pathologists and proper patho-clinical and neuro-oncological cooperation" in *Folia Neuropathologica*, 60(2), 137-152, 2022.
- [2] A. Gautheron "Towards quantitative multispectral excitation fluorescence spectroscopy for intraoperative assistance in glioma resection in neurosurgery" *Biotechnology*. Université Claude Bernard - Lyon I, 2022
- [3] A. Gautheron et al., "Experimental Decorrelation Requirement for Robust Estimation of Fluorophores Using Multiple-Wavelength Excitation Fluorescence Spectroscopy", in *Proceedings Volume 12628, Diffuse Optical Spectroscopy and Imaging IX; 126282F*, 2023
- [4] P. Leclerc, C. Ray, L. Mahieu-Williams et al, "Machine learning-based prediction of glioma margin from 5-ALA induced PpIX fluorescence spectroscopy" in *Scientific Reports* 10, 1462, 2020.
- [5] T. Picart, A. Gautheron, C. Caredda, et al. "Fluorescence-Guided Surgical Techniques in Adult Diffuse Low-Grade Gliomas: State-of-the-Art and Emerging Techniques: A Systematic Review" in *Cancers (Basel)* ;16(15):2698, 2024.
- [6] T. Savelieva, I. Romanishkin, A. Ospanov, S. Goryaynov, G. Pavlova, I. Pronin, V. Loschenov "Machine Learning and Artificial Intelligence Systems Based on the Optical Spectral Analysis in Neuro-Oncology" in *Photonics*, 12(1), 37, 2025.
- [7] H. Abdi, L. J. Williams "Principal component analysis" in *Wiley interdisciplinary reviews: computational statistics* 2.4, 433-459, 2010
- [8] P. Xanthopoulos et al. "Linear discriminant analysis" in *Robust data mining*: 27-33, 2013
- [9] N. V. Chawla et al. « SMOTE: Synthetic Minority Over-sampling Technique" in *Journal of Artificial Intelligence Research* 16 321–357, 2002