



## **Machine learning-based prediction of glioma margin from 5-ALA induced PpIX fluorescence spectroscopy**

Pierre Leclerc, Cédric Ray, Laurent Mahieu-Williame, Laure Alston, Carole Frindel, Pierre-Francois Brevet, David Meyronet, Jacques Guyotat, Bruno Montcel, David Rousseau

### **► To cite this version:**

Pierre Leclerc, Cédric Ray, Laurent Mahieu-Williame, Laure Alston, Carole Frindel, et al.. Machine learning-based prediction of glioma margin from 5-ALA induced PpIX fluorescence spectroscopy. Scientific Reports, 2020, 10 (1), pp.1462 et suiv. <10.1038/s41598-020-58299-7>. <hal-02440653v2>

**HAL Id: hal-02440653**

**<https://hal.science/hal-02440653v2>**

Submitted on 29 Jan 2020

**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

OPEN

# Machine learning-based prediction of glioma margin from 5-ALA induced PpIX fluorescence spectroscopy

Pierre Leclerc<sup>1,2</sup>, Cedric Ray<sup>1</sup>, Laurent Mahieu-William<sup>2</sup>, Laure Alston<sup>2</sup>, Carole Frindel<sup>2</sup>, Pierre-François Brevet<sup>1</sup>, David Meyronet<sup>3,4</sup>, Jacques Guyotat<sup>3</sup>, Bruno Montcel<sup>2,6\*</sup> & David Rousseau<sup>2,5,6</sup>

Gliomas are infiltrative brain tumors with a margin difficult to identify. 5-ALA induced PpIX fluorescence measurements are a clinical standard, but expert-based classification models still lack sensitivity and specificity. Here a fully automatic clustering method is proposed to discriminate glioma margin. This is obtained from spectroscopic fluorescent measurements acquired with a recently introduced intraoperative set up. We describe a data-driven selection of best spectral features and show how this improves results of margin prediction from healthy tissue by comparison with the standard biomarker-based prediction. This pilot study based on 10 patients and 50 samples shows promising results with a best performance of 77% of accuracy in healthy tissue prediction from margin tissue.

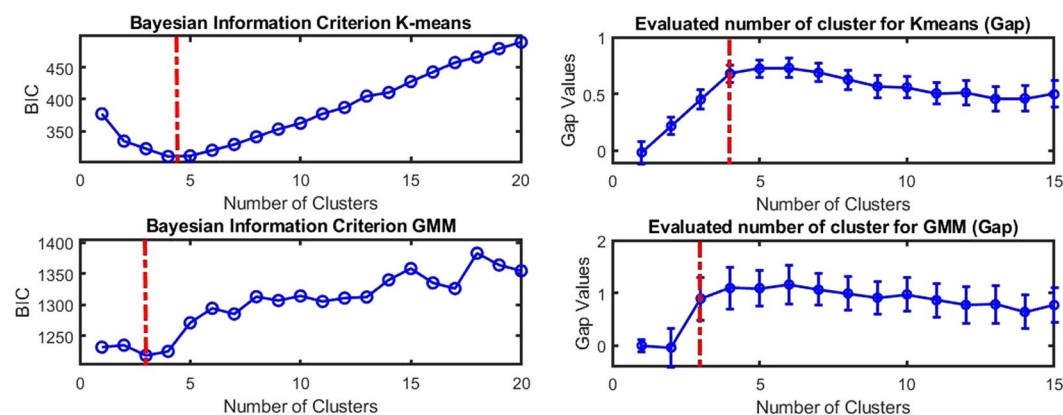
Gliomas account for more than fifty percent of primitive brain tumors. They are infiltrative tumors, with a margin difficult to identify and discriminate from the surrounding healthy tissues. The world health organization (WHO) classifies gliomas in 4 grades<sup>1</sup>, but most studies commonly consider two separate groups: High-Grade Gliomas (HGG) and Low-Grade Gliomas (LGG). Studies have shown that in 85% cases, recurrences of HGG are localized less than 2 centimeters away from the initial tumor<sup>2</sup>. Then, improving the extent of resection is relevant to prevent recurrence and improve life quality and expectancy<sup>3–5</sup>. Pre-operative MRI combined with neuro-navigation is currently used in the operating theater<sup>6,7</sup> but shows strong limitations<sup>8–10</sup>. 5-aminolevulinic acid (5-ALA) induced protoporphyrin IX (PpIX) fluorescence microscopy has shown its relevance in neuro-oncology<sup>11</sup>. PpIX absorbs light at 405 nm and emits fluorescence with a main peak centered at 634 nm. This technique is the actual clinical standard for PpIX-based surgical assistance. However, its sensitivity is still limited when applied to low-density infiltrative parts of HGG<sup>12,13</sup> or to LGG<sup>14</sup>.

Various 5-ALA induce PpIX fluorescence spectroscopy methods have been proposed to overcome these sensitivity issues. Previous works<sup>6,15–24</sup>, focus on the extraction of biomarkers from the measurements, based on a priori information on the link between the biomarkers and the microenvironment of PpIX. These approaches are known as expert-based, and various biomarker models have been proposed in the literature. Quantification of PpIX concentration<sup>15</sup> show enhanced sensitivity either in HGG<sup>16</sup> or in LGG<sup>17</sup>. Normalization procedures of biomarkers can also increase their robustness<sup>6,18,19</sup>. Other works suggest that relevant models could be obtained based on the shape of the PpIX emission spectrum<sup>18–26</sup>. These works show that the PpIX fluorescence emission spectral complexity in tissue is closely linked with the pathological status. However, the still unsolved origin of this complexity impairs the extraction of the best features with an expert-based related method, thus preventing the classification of measurements into relevant pathological status.

<sup>1</sup>Univ Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-69622, Villeurbanne, France, 10 Rue Ada Byron, 69622, Villeurbanne, France. <sup>2</sup>CREATIS, Univ Lyon, CNRS UMR5220, INSERM U1044, Université Claude Bernard Lyon1, INSA Lyon, Villeurbanne, France. <sup>3</sup>Hospices Civils de Lyon, Centre de Pathologie et de Neuropathologie Est, Lyon, F-69003, France. <sup>4</sup>Cancer Research Centre of Lyon, Univ Lyon, INSERM U1052, CNRS UMR5286, Lyon, France, Université Claude Bernard Lyon 1, Lyon, France. <sup>5</sup>Laboratoire Angevin de Recherche en Ingénierie des Systèmes, UMR INRA IRHS, Université d'Angers, 62 avenue Notre Dame du Lac, 49000, Angers, France. <sup>6</sup>These authors contributed equally: Bruno Montcel and David Rousseau. \*email: [bruno.montcel@univ-lyon1.fr](mailto:bruno.montcel@univ-lyon1.fr)



**Figure 1.** Global view of the proposed machine learning-based prediction of glioma margin by PpIX fluorescence spectroscopic measurements. In this study, the data set is composed of 50 samples from 10 patients. From left to right, the optical spectrum of cells around a tumor is measured. The dimension of the spectral information is then reduced to lower the redundancy. Supervised or unsupervised algorithms are finally used to classify the data and create a prediction of tissue state from the PpIX fluorescence spectroscopic measurements.



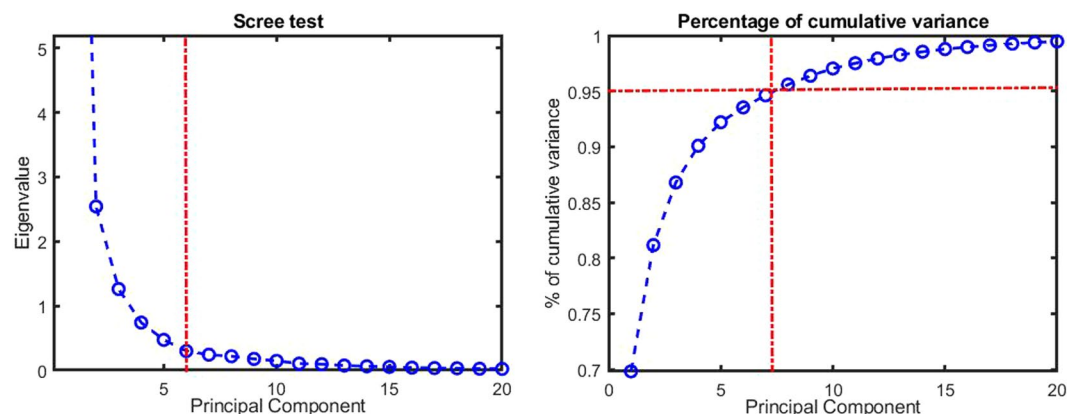
**Figure 2.** Bayesian inference criterion (BIC) (left) and gap criterion (right) as a function of the number of clusters for K-means (top row) and GMM (bottom row). The minimum of the BIC and maximum of the gap criterion (highlighted in the red dash-dotted line) correspond to the optimal number of clusters in our data. Interestingly K-means and GMM are best described with 3 or 4 clusters which fit with the red dotted lines corresponding to the number of classes from the clinical taxonomy.

In this study, we adopt a different approach for the prediction of glioma margin from fluorescence information. Instead of choosing a small amount of numerical features used as biomarkers like in the recent above-cited literature<sup>6,15–24</sup>, we propose to investigate the prediction of glioma margin with an entirely data-driven approach where no assumption of feature selection based on expert is made. To this purpose, we implement, for the first time to our knowledge in this context, a machine learning classification approach. The pipeline, as illustrated in Fig. 1 and detailed in the material and method section predicts glioma margin from the raw fluorescence spectrum. This is done on the same data acquired previously in a surgical procedure of 5-ALA induced PpIX guided glioma removal and which had been only processed so far with a biomarker approach<sup>23</sup>. This choice enables a comparison of the prediction performance of glioma margin from a biomarker approach with a novel expert-independent point of view. Also, as another element of novelty, the prediction of glioma margin is performed from 3 different fluorescent spectra in response to 3 distinct excitation wavelengths taken successively over the same area, while previously only a few features extracted from a single fluorescent spectrum were used for analysis<sup>23</sup>.

## Results

Before comparing performances of supervised and unsupervised classification of machine-learning-based with an expert-based approach, we provide the estimation of the number of clusters and the identification of the best spectral features with the pipeline of Fig. 1 as described step-by-step in the Method section.

**Data driven estimation of the number of clusters corresponds to the clinical taxonomy.** The number of classes in the data set was automatically estimated using data-driven methods. As illustrated in Fig. 2, the optimum number of clusters is robustly found between 4 and 5. This is recorded by the Bayesian information criterion (BIC)<sup>27</sup> and gap criteria<sup>28</sup> where the extremum of the curves indicate the optimal number of clusters for both tested clustering methods (K-Means, Gaussian Mixture Models). Interestingly, while obtained here from a purely data-driven approach, this number of clusters is compatible (see red dotted lines in Fig. 2) with the number of classes proposed independently by the clinical taxonomy described in<sup>23</sup> from histological images: tumor core, high-density margin, low-density margin, healthy tissue.



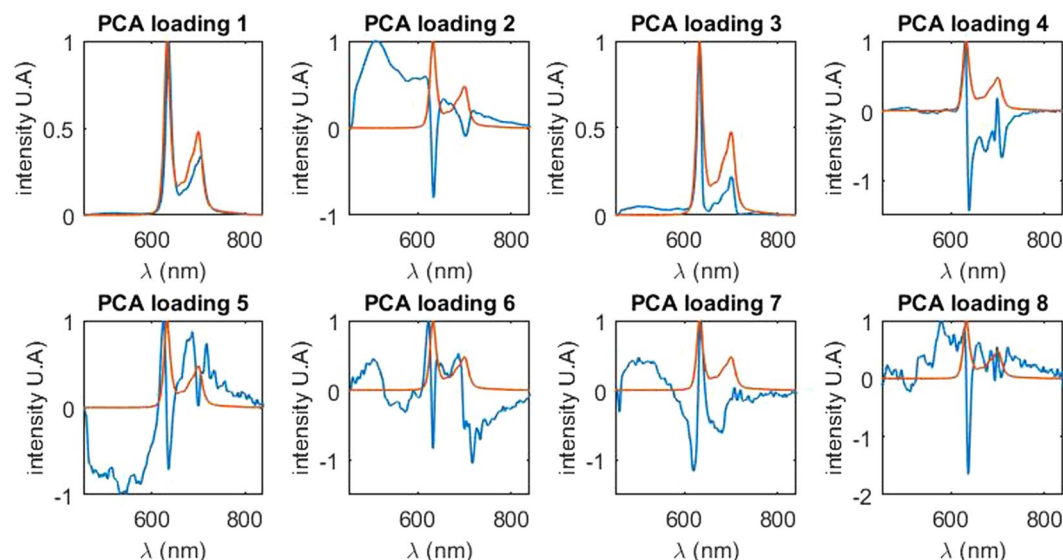
**Figure 3.** Scree test (on the left) and cumulative variance (on the right) of principal component analysis. A minimum of 5 principal components is required to describe the data variance as can be inferred from the Scree test “elbow” and the saturation around 95% of the cumulative variance highlighted in red dotted lines.

**Identification of best spectral features.** Our original feature space composed of raw fluorescence spectrum with 900 emission wavelengths (from 435 to 840 nm) taken at three distinct excitation wavelengths includes 2700 features. In order to both select the best spectral features and analyze their shapes, dimension reduction techniques were applied. Based on the principal component analysis (PCA)<sup>29</sup>, the Scree test<sup>30</sup> and the cumulative variance were used to find the statistically relevant principal spectral components. As illustrated in Fig. 3, either the Scree test (left) and the cumulative variance (right) in the PCA analysis show that a minimum of 5 components is required to describe the data variance. Indeed, the Scree test “elbow” and the saturation around 95% of cumulative variance are both found around five components (given by the red dotted vertical line in Fig. 3). By comparison, in our previous study<sup>23</sup>, for one excitation wavelength, 2 features were extracted from the spectrum: the relative intensity of the component leading to a peak of PpIX fluorescence at 620 nm (PpIX620) and the relative intensity of the component leading to a peak PpIX of fluorescence at 634 nm (PpIX634). With three excitation wavelengths, the resulting feature space of this model is  $2 \times 3 = 6$  descriptors to describe the variability of the data. Thus the number of statistically relevant descriptors computed with the data-driven approach with the principal components analysis (5 components) is compatible with the expert-based model (6 components). In addition to the number of relevant principal components, the shapes of these components were analyzed in the original feature space, *i.e.* the fluorescence emission spectrum. These are plotted in Fig. 4. In order to simplify the visualization, only the result for one excitation wavelength (405 nm) is displayed. However, results were very similar to the two other excitation wavelengths.

It is then interesting to compare the shape of the principal components extracted from the data-driven PCA approach with the spectral patterns extracted in the expert-based approach of<sup>23</sup>, as shown in Fig. 4. Not surprisingly, the first principal component is very similar to the fluorescence emission spectrum of the PpIX<sup>19</sup> with a peak around 634 nm and, also not surprisingly the second component corresponds to the autofluorescence of the tissue, which can be linked to numerous endogenous autofluorescent molecule such as NADH, FAD, lipofuscin<sup>31,32</sup>. These molecules are also known to be correlated with cancerous pathological status through the Warburg effect<sup>22,33</sup> which modify cell metabolism and favorize glycolysis. The third principal component is a mix between the spectrum of PpIX with a peak around 632 nm and the fluorescence previously described. The fourth principal component is comprised of a peak at 629 nm and a smaller one at 695 nm. The fifth and last relevant principal component shows a maximum at 625 nm. All principal components present a peak varying from 622 nm (principal component 6) to 636 nm (principal component 1). Interestingly, the variance of our normalized spectra is best described as a blue-shift of the PpIX spectrum peak from 636 nm to 620 nm. Remarkably, the shape of these principal components extracted with a purely data-driven approach corresponds to spectral components identified in the expert-based studies<sup>19,23</sup>.

**Good predictions are obtained in unsupervised and supervised modes.** Unsupervised learning was then used to find clusters among a reduced feature space autonomously. Two methods were tested, including PCA and T-SNE algorithms<sup>34</sup>. A dimension reduction to 3 was chosen to facilitate the visualization of clusters. With this choice, the resulting clustering presented a better accuracy with T-SNE rather than PCA, and thus only T-SNE results are presented here when applied to K-means and GMM for the four classes identified previously. As K-Means and GMM are randomly initiated, in order to assess the quality of our results, the classification was reproduced twenty times and averaged. Results can be seen for four classes with K-means in Table 1 when the entire feature space (3 excitation wavelengths) was considered. A comparison with a feature space based on each excitation wavelength was also plotted and discussed further in the Supplementary Material. The typical clustering results are displayed in Fig. 5.

As seen in Fig. 5, the four resulting computed clusters fairly correspond to the clinical taxonomy given by the anatomopathologist (on the left). In addition, these clusters are in this feature space aligned along with the ordinal severity of symptoms following the density of diseased cells with maximization of the distance between tumor core and healthy cells. Both K-means and GMM methods (middle and right of Fig. 5, respectively) give results



**Figure 4.** Normalized principal component of the PCA (in blue) in the original feature space (*i.e.* the optical spectrum space). For each principal component, the reference spectrum of the PpIX (for the state peaking at 634 nm) is also plotted (in red). For better comprehension, only one of the three fluorescence emission spectrum from the original feature space is represented as they are similar for all three excitation wavelengths. The first principal component is similar to the spectrum of the PpIX with a peak of 636 nm. The second principal component is best described as the autofluorescence of the measured tissue, *i.e.* the contribution of other fluorophores. The five following components all show a peak shifting between 620 nm and 636 nm.

	Predicted Core	Predicted HD Margin	Predicted LD Margin	Predicted Healthy
True Core (10)	8 (80%) $\sigma = 0$	1.9 (19%) $\sigma = 0.31$	0 (0%) $\sigma = 0$	0.1 (1%) $\sigma = 0.31$
True HD Margin (24)	1.3 (5%) $\sigma = 0.5$	7.1 (30%) $\sigma = 0.32$	9.7 (40%) $\sigma = 0.48$	5.9 (25%) $\sigma = 0.32$
True LD Margin (9)	0 (0%) $\sigma = 0$	0.1 (1%) $\sigma = 0.31$	5.6 (62%) $\sigma = 0.52$	3.3 (37%) $\sigma = 0.48$
True Healthy (7)	0 (0%) $\sigma = 0$	0 (0%) $\sigma = 0$	1 (14%) $\sigma = 0$	6 (86%) $\sigma = 0$

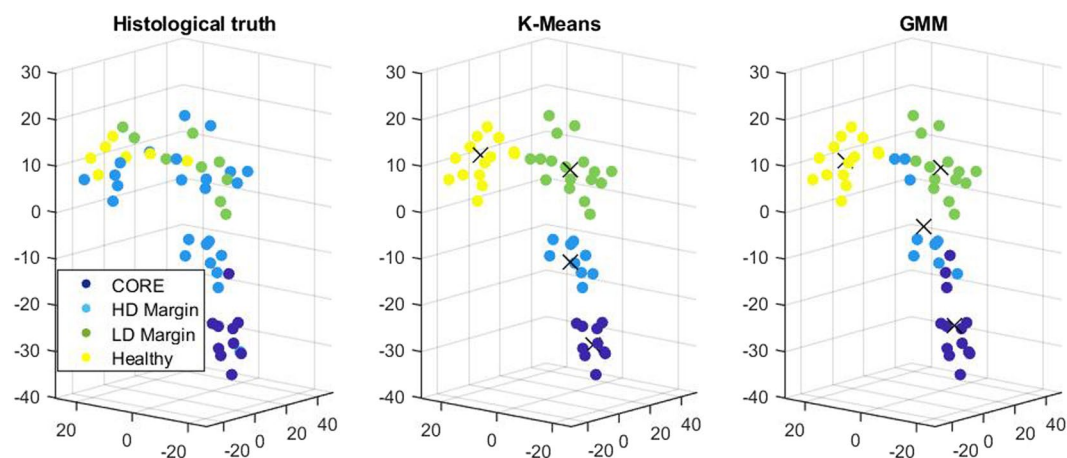
**Table 1.** Confusion matrix for K-means with 4 classes: tumor core, high density margin, low density margin and healthy tissue. Statistics (average, standard deviation) result from 20 predictions. In each cell, from left to right: number of instances, percentage of total class population and standard deviation.

in accordance with the histological truth. Figure 5 and Table 1 constitutes promising results. While accurate prediction for the tumor core is logical since it is the main target of the PpIX, we also see that accurate prediction is obtained for the healthy tissue. Overall a prediction between healthy tissue, tumor core, and margin led to a 73% accuracy and went up to 77% for classification with only three classes. To further stress the interest of machine learning-based prediction of glioma margin, a supervised method was also used. Typical results are shown in the Supplementary Material.

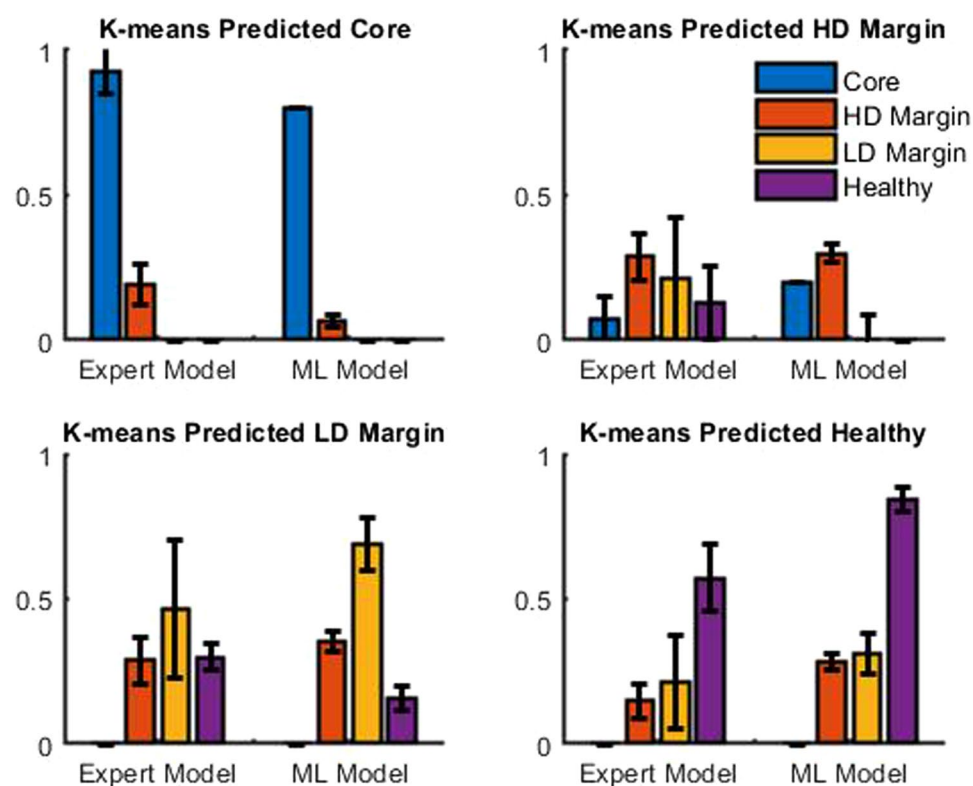
**Comparison between machine-learning-based and expert-based feature selection.** As stated above, the number of clusters and the shape of the principal component calculated are remarkably similar to what has been described in<sup>23</sup>. The benefits of this fully automated approach compared to the expert-based model is thus investigated. We use the expert-based method of our previous study<sup>23</sup>. This consist in a fitting process of the fluorescence spectrum to retrieve the relative contribution of the two states of PpIX, PpIX620 and PpIX634 for each excitation wavelength. In this study this process is led for the three excitation wavelength, instead of only at the 405 nm excitation wavelength in the previous study. The resulting feature space of dimension 6 is then normalized and reduced using T-SNE to 3 dimensions like with the machine learning approach in order to enable a fair comparison between the discrimination power of each method. The average result of 20 predictions can be seen in Fig. 6. The machine learning-based method shows a clear superiority over the expert-based model. In particular, the “Healthy tissue” case is significantly better for true positive and true negative while being comparable or better for false positive.

This result demonstrates that despite similar features, the fully automated machine-learning-based method shows significantly better results than the previously used expert-based model. In the expert-based model of<sup>23</sup>, the fit was between 585 and 640 nm, which did not include patterns around 700 nm which shows variability among the datasets. Also, in<sup>23</sup> the autofluorescence was used to normalize every spectrum and was then subtracted from the signal. In our case, autofluorescence was used differently since it was kept in the signal (mainly present in the second component of the PCA in Fig. 4).





**Figure 5.** Unsupervised classification in a T-SNE reduced space. From left to right: the histological truth as given by the anatomopathologist, K-means classification and GMM classification both with 4 clusters.



**Figure 6.** Comparison of confusion matrix for K-means with 4 classes: tumor core, high and low density margin and healthy tissue. Average result of 20 predictions. ML model is Machine learning-based Model, HD stands for high density and LD for low density.

## Discussion

This study demonstrated the interest of a machine learning-based approach for the prediction of glioma margin. This approach found spectral features important for the prediction of glioma margin, which happens to be close to the one selected in the expert-based model of<sup>23</sup>. This is an important result since this is obtained by two independent ways on the same dataset. It reinforced, quantitatively, the evidence of the interest in using the blue-shift in the PpIX fluorescence spectrum as a mean of discrimination between margin and healthy tissue. Other works already suggested a wavelength blue-shift of the peak intensity of the emission spectrum correlated with the tissue pathological status<sup>18,19,23</sup>. In particular, a second peak of fluorescence at 620 nm has been observed in tissues<sup>19-23,25,26</sup> or in cell culture<sup>21,24</sup>. It is *in vivo* origin is still an open issue, some works supporting the assumption that it is related to different aggregates of PpIX<sup>19,23,24,26</sup>, other works<sup>20,35</sup> argue that it is a fluorescence induced

by the precursors of PpIX, uroporphyrins or coproporphyrins. This work cannot give clues on the origin of the blue-shift. However, the relevance of the blue-shift effect is retrieved in this data-driven approach. This reinforces the expert-driven previous works<sup>18–26</sup> focusing on the blue-shift. The biological mechanism behind this shift remains to be uncovered to this day and is not the subject of this work. In this section, we rather discuss some elements of data preparation, excitation wavelengths, and further machine learning approaches.

**Data sample preparation.** By contrast with<sup>23</sup>, the distinction between LGG and HGG was not done to increase the sample size of clusters. This choice could be discussed since the pathology is classified as different in high grade and low grade from the perspective of histology, which is based on tissular structures at the supra-cellular scale. However, a posteriori, we found that this choice does not prevent a reasonably good clustering prediction for the whole data set. In a future study, with increased sample size, it would be interesting to compare the clustering with labels differentiating LGG and HGG and the fused approach followed in this study.

**Multi-excitation wavelengths.** The guideline of the article is to investigate how the increase of the fluorescence feature space can contribute to improving the performance of classification by comparison with an approach based on few fitted parameters. We extend the feature space to the entire spectrum of a single excitation wavelength in the core of the manuscript. We demonstrated that this feature space extension produces a gain of classification performances. In the complementary material of the article, we also report the performance when the feature space is increased to an additional fluorescence spectrum obtained with other excitation wavelengths. In order to probe the two PpIX states (peaking at 634 nm and 620 nm), the optical fluorescence spectra were acquired for three different LEDs exposition: 385, 405 and 420 nm. We investigated the potential of using different excitation wavelengths to expand the feature space, suppress potential degeneracy, and potentially get more accurate information and better insight. Using a PCA, we analysed the principal component in the original feature space (the eigenvectors in the spectral space). However, the resulting PCA were rather similar for all 3 LEDs, not showing significantly different information with respect to the wavelength. This is probably due to the significant overlapping of the emission spectra of the 3 LEDs. Another explanation for this degeneracy can be that minor change in the spectrum acquired with different excitation wavelengths are not distinguished because the small sample size does not allow for a minor and subtle change in the spectrum to be discriminating in this study.

**Supervised learning.** Supervised learning showed (in Supplementary Material Section) consistent results with unsupervised learning. This is an interesting result since it shows the robustness of the classification with various machine learning algorithms. While it is certainly encouraging, we cannot be definitive about these supervised methods until the sample size is increased. Such input of more data would allow a fine-tuning of these models or a more advanced supervised classifier (such as deep neural networks), including more hyperparameters. Extended cohort and larger training dataset would also enable to perform classification in real time for clinical use without the need to perform cross-validation.

## Conclusion

In this article, we have demonstrated the interest of a machine learning approach for the prediction of glioma margin from 5-ALA induced PpIX fluorescence spectroscopy. When considering the entire raw spectrum as input feature space, classical dimension reduction was shown to select spectral patterns similar to those identified around 634 nm and 620 nm as possible biomarkers for margin prediction in our previous expert based work<sup>23</sup>. This pure data-driven proof, independent from the expert-based approach found in the current literature, is significant since the biochemical or physical origin of these spectral biomarkers is not yet understood. A second interest of the machine learning approach proposed here is that it shows an increase of discrimination as compared to our previous expert-based features used as biomarkers<sup>23</sup>. The best performance of 77% of accuracy between healthy tissue and margin is found. Despite the relatively small size of the data set considered here, this can be considered as promising pilot results due to their self-consistency with the classical expert-based approach for feature selection. Repetition with a larger cohort will have to be carried out to establish the added value of the optical probe for surgery. With such an extended data set, other machine learning models with higher capability and tuning parameters could be tested. Other expert model<sup>16,17</sup> performance could also be compared against data driven approach to test its robustness. Also, another direction of investigation for the future could be to enlarge the feature space. No positive effect on prediction performance was recorded when increasing the number of excitation with single-photon fluorescence in this study. Two-photons fluorescence or the effect of polarization could also be worth to investigate in this context while revisiting the proposed machine-learning approach.

## Methods

The data set was acquired during a clinical study led at the neurologic center of the Pierre Wertheimer hospital in Bron, France. This study was described in detail in previous works<sup>23</sup>. 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 signed written informed consent. Only a part of the acquired data was used since we focused on “*in vivo*” measurements and used the multi-wavelength excitation measurements. For readability purposes, we shortly described the method and added complementary information to the already published method of<sup>23</sup>.

**Spectroscopic device.** The developed device has been described in detail in previous works<sup>23</sup>. Here, as a novelty, multi-wavelengths excitation, not described previously, was used. Therefore, we describe here the setup, including the multi-wavelength excitation capabilities. Excitation was performed through three light-emitting

diodes (LED) centered at 385 nm, 405 nm, and 420 nm with 7 nm Full-Width Half-Maximum (M385F1, M405F1, M420F1, Thorlabs). Emitted light was transmitted through three optical fibers (HCG M0600T, sedi-fibres) to a dedicated probe. The probe entrance consists of a bundle of 7 optical fibers of 600  $\mu\text{m}$  core diameter. The other ends of these fibers are cleaved, so that excited tissue area and emitting tissue area are the same. The light goes through a low-pass filter (Edmunds Optics OD4 low pass 450 nm) with a cutoff wavelength of 450 nm. This led to an output irradiance of 80 W/m, 30 W/m, and 50 W/m, respectively, for the LED centered at 385 nm, 405 nm, and 420 nm. Tissue reflectance was collected through the same probe, with a detection fiber, and goes through a high pass filter (HQ485LP, Chroma) with a cutoff wavelength of 485 nm. The filtered light was finally injected into a spectrometer (Maya2000, Ocean optics). Characterization of the system has been performed on calibrated phantoms<sup>36</sup>.

**Surgical procedure and data acquisition.** Patients were given an oral dose of 20 mg/kg of body weight of 5 aminolevulinic acids (Gliolan; Medac GmBH) approximately 3 hours prior to the induction of anesthesia. For each patient, the standard surgical procedure started in order to expose the tissue, and, when asked by the surgeon, the surgical procedure was stopped so that fluorescence measurements were performed. Each acquisition was composed of 200 ms of duration, with the LED turned on, followed by the same duration with the LED turned off to get rid of ambient light coming from the operating room. For each measurement, 6, 12, and 6 acquisitions were led respectively for the LED centered at 385 nm, 405 nm, and 420 nm. This gave a total acquisition time of 9.6 s. The tissue was then removed and sent for histopathological analysis. These fluorescence measurements were performed in order to get different densities of infiltrative tumor cells per glioma. In total, 50 measurements were kept in this analysis.

**Histopathology.** Histopathological analysis was performed on formalin-fixed paraffin-embedded biopsy tissue specimens processed for H & E staining. Each H & E stained tissue section was assessed for the presence of tumor cells, necrosis, mitotic activity, nuclear atypia, microvascular proliferation, and reactive astrogliosis. Molecular criteria were also assessed. Biopsy specimens were then classified into five categories based on WHO histopathological and molecular criteria<sup>1</sup> as HGG solid part, HGG margin, HGG margin of low density, LGG and healthy tissue. Finally, LGG and HGG data were combined in order to increase the sample size and test the expert driven against data driven approaches. This clustering is supported by previous works<sup>19</sup> showing that LGG and HGG margins share common properties in terms of PpIX fluorescence intensity. The samples from LGG patients were included in the healthy, HGG low-density margin, or HGG high-density margin depending on their pathological status. The resulting studied data set was composed of 50 samples from 10 patients. This includes 28 samples for HGG composed of 10 samples from tumor core, eight from the high-density margin, five from the low-density margin, and five healthy samples. Furthermore, this includes also 22 samples for LGG, 17 included in HGG high-density margin, three included in HGG low-density margin, and two included in healthy tissue, depending on their pathological status.

**Data processing pipeline.** The data processing pipeline developed is illustrated in Fig. 1. In the first step, the spectrum was acquired with the optical system. Three fluorescence spectra corresponding to three excitation LED's were acquired. The pathological status of the corresponding tissue was recorded from the histopathological analysis. In the second step, the spectrum was then normalized by the global energy of the spectrum. The feature spaces, created by the three spectra, being too large to be applied to clustering methods<sup>37,38</sup> were reduced with dimension reduction techniques in a third step. In our case, principal component analysis (PCA) and t-distributed stochastic neighbor embedding (T-SNE), which are two basic techniques for dimension reduction, were tested<sup>34,39</sup>. PCA was mainly chosen among others, orthogonal transformations, for its exploratory ability to summarize data along with their main characteristics. T-SNE was mainly chosen for its ability to preserve local structure so that points close to one another in the high-dimensional feature space will tend to be close to one another in the reduced feature space. The number of principal components to retain is determined with the Scree test<sup>30</sup> or by computing the number of principal components requested to reach 95% of cumulated variance. The last step corresponded to the classification from the reduced feature space in an unsupervised and supervised way. Two clustering methods were chosen for this study: K-Means<sup>40</sup> and the Gaussian mixture model (GMM)<sup>41</sup>. A large variety of methods can be found in the literature with various levels of complexity and hyperparameters to be adjusted<sup>42</sup>. Because the size of the data set is fairly small, algorithms were chosen with a minimal amount of hyperparameters to be adjusted. K-Means cluster points in the feature space inside hyper-spheres according to a euclidian distance while GMM includes an additional degree of freedom in the organization of the points which are clustered inside hyper-ellipsoids. In K-Means and GMM, the number of clusters is a hyperparameter, which was determined with the Bayesian inference criterion<sup>27</sup> and the "gap" criteria<sup>28</sup>. We compared the predictive value of the pure data-driven feature space with the predictive value of expert-based feature space. The expert-based feature space consists of two features  $\alpha_{634}$  and  $\alpha_{620}$  computed from the raw spectrum. As described in<sup>23</sup>, the other intrinsic fluorophores (at the exclusion of PpIX) emitting below 620 nm were removed in the emission spectrum. The resulting spectrum was then fitted with the contribution of two PpIX spectra acquired *in vitro*. One of those spectrums presented a peak assumed Gaussian at 620 nm, and its relative contribution to the resulting spectrum was named  $\alpha_{620}$ . The other one presented a peak also assumed Gaussian at 634 nm, and its relative contribution was named  $\alpha_{634}$ . An extensively detailed explanation of this feature space can be found in previous work<sup>23</sup>. The reference spectra in Fig. 4 was also taken from this previous work.

## Data availability

The datasets analysed during the current study are available from the corresponding author on reasonable request.



Received: 5 July 2019; Accepted: 6 January 2020;

Published online: 29 January 2020

## References

- Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* **131**, 803–820, <https://doi.org/10.1007/s00401-016-1545-1> (2016).
- Petrecca, K., Guiot, M.-C., Panet-Raymond, V. & Souhami, L. Failure pattern following complete resection plus radiotherapy and temozolomide is at the resection margin in patients with glioblastoma. *J. Neuro-Oncology* **111**, 19–23, <https://doi.org/10.1007/s11060-012-0983-4> (2013).
- Laws, E. R. *et al.* Survival following surgery and prognostic factors for recently diagnosed malignant glioma: data from the Glioma Outcomes Project. *J. Neurosurg.* **99**, 467–473, <https://doi.org/10.3171/jns.2003.99.3.0467> (2003).
- Lacroix, M. & Toms, S. A. Maximum Safe Resection of Glioblastoma Multiforme. *J. Clin. Oncol.* **32**, 727–728, <https://doi.org/10.1200/JCO.2013.53.2788> (2014).
- Li, Y. M., Suki, D., Hess, K. & Sawaya, R. The influence of maximum safe resection of glioblastoma on survival in 1229 patients: Can we do better than gross-total resection? *J. Neurosurg.* **124**, 977–988, <https://doi.org/10.3171/2015.5.JNS142087> (2016).
- Haj, A. *et al.* Extent of resection in newly diagnosed glioblastoma: Impact of a specialized neuro-oncology care center. *Brain Sci.* **8**, 5, <https://doi.org/10.3390/brainsci8010005> (2018).
- Wu, G. N., Ford, J. M. & Alger, J. R. MRI measurement of the uptake and retention of motexafin gadolinium in glioblastoma multiforme and uninvolved normal human brain. *J. Neuro-Oncology* **77**, 95–103, <https://doi.org/10.1007/s11060-005-9101-1> (2006).
- Nabavi, A. *et al.* Serial Intraoperative Magnetic Resonance Imaging of Brain Shift. *Neurosurgery* **48**, 787–798, <https://doi.org/10.1097/00006123-200104000-00019> (2001).
- Kubben, P. L. *et al.* Intraoperative MRI-guided resection of glioblastoma multiforme: a systematic review. *The Lancet Oncol.* **12**, 1062–1070, [https://doi.org/10.1016/S1470-2045\(11\)70130-9](https://doi.org/10.1016/S1470-2045(11)70130-9) (2011).
- Senft, C. *et al.* Intraoperative MRI guidance and extent of resection in glioma surgery: a randomised, controlled trial. *The Lancet Oncol.* **12**, 997–1003, [https://doi.org/10.1016/S1470-2045\(11\)70196-6](https://doi.org/10.1016/S1470-2045(11)70196-6) (2011).
- Stummer, W. *et al.* Technical Principles for Protoporphyrin-IX-Fluorescence Guided Microsurgical Resection of Malignant Glioma Tissue. *Acta Neurochir.* **140**, 995–1000, <https://doi.org/10.1007/s007010050206> (1998).
- Bravo, J. J. *et al.* Hyperspectral data processing improves PpIX contrast during fluorescence guided surgery of human brain tumors. *Sci. Reports* **7**, 9455, <https://doi.org/10.1038/s41598-017-09727-8> (2017).
- Hadjipanayis, C. G., Widhalm, G. & Stummer, W. What is the surgical benefit of utilizing 5-Aminolevulinic acid for fluorescence-guided surgery of malignant gliomas? *Neurosurgery* **77**, 663–673, <https://doi.org/10.1227/NEU.0000000000000929> (2015).
- Jaber, M. *et al.* Is visible aminolevulinic acid-induced fluorescence an independent biomarker for prognosis in histologically confirmed (World Health Organization 2016) low-grade gliomas? *Neurosurgery* **84**, 1214–1224, <https://doi.org/10.1093/neuros/nyy365> (2019).
- Valdés, P. A. *et al.* Quantitative, spectrally-resolved intraoperative fluorescence imaging. *Sci. Reports* **2**, 798, <https://doi.org/10.1038/srep00798> (2012).
- Valdés, P. A. *et al.* System and methods for wide-field quantitative fluorescence imaging during neurosurgery. *Opt. Lett.* **38**, 2786–2788, <https://doi.org/10.1364/OL.38.002786> (2013).
- Valdés, P. A. *et al.* Quantitative fluorescence using 5-aminolevulinic acid-induced protoporphyrin IX biomarker as a surgical adjunct in low-grade glioma surgery. *J. Neurosurg.* **123**, 771–780, <https://doi.org/10.3171/2014.12.JNS14391> (2015).
- Ando, T. *et al.* Precise comparison of protoporphyrin IX fluorescence spectra with pathological results for brain tumor tissue identification. *Brain Tumor Pathol.* **28**, 43–51, <https://doi.org/10.1007/s10014-010-0002-4> (2011).
- Montcel, B., Mahieu-Williams, L., Armoiry, X., Meyronet, D. & Guyotat, J. Two-peaked 5-ALA-induced PpIX fluorescence emission spectrum distinguishes glioblastomas from low grade gliomas and infiltrative component of glioblastomas. *Biomed. Opt. Express* **4**, 548–558, <https://doi.org/10.1364/BOE.4.000548> (2013).
- Zanella, M. *et al.* Multimodal optical analysis discriminates freshly extracted human sample of gliomas, metastases and meningiomas from their appropriate controls. *Sci. Reports* **7**, 41724, <https://doi.org/10.1038/srep41724> (2017).
- Dietel, W., Fritsch, C., Pottier, R. H. & Wendenburg, R. 5-aminolevulinic-acid-induced formation of different porphyrins and their photomodifications. *Lasers Med. Sci.* **12**, 226–236, <https://doi.org/10.1007/BF02765103> (1997).
- Figueredo, A. J. & Wolf, P. S. A. Assortative pairing and life history strategy. *Hum. Nat.* **20**, 317–330, <https://doi.org/10.1007/s12110-009-9068-2> (2009).
- Alston, L. *et al.* Spectral complexity of 5-ala induced ppx fluorescence in guided surgery: a clinical study towards the discrimination of healthy tissue and margin boundaries in high and low grade gliomas. *Biomed. Opt. Express* **10**, 2478–2492, <https://doi.org/10.1364/BOE.10.002478> (2019).
- Hope, C. K. & Higham, S. M. Evaluating the effect of local pH on fluorescence emissions from oral bacteria of the genus *Prevotella*. *J. Biomed. Opt.* **21**, 084003, <https://doi.org/10.1117/1.JBO.21.8.084003> (2016).
- Montcel, B. *et al.* 5-ALA-induced PpIX fluorescence in gliomas resection: spectral complexity of the emission spectrum in the infiltrative compound. In *SPIE Proceedings*, 8804, 880409 (Optical Society of America, 2013).
- Alston, L. M. *et al.* Interventional fluorescence spectroscopy: preliminary results to detect tumor margins during glioma resection with two fluorescence spectra of PpIX. In *SPIE Proc.* 10411, 104110C, <https://doi.org/10.1117/12.2286110> (Optical Society of America, 2017).
- Schwarz, G. Estimating the dimension of a model. *The Annals Stat.* **6**, 461–464, <https://doi.org/10.1214/aos/1176344136> (1978).
- Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. Royal Stat. Soc. Ser. B (Statistical Methodol.)* **63**, 411–423, <https://doi.org/10.1111/1467-9868.00293> (2001).
- Pearson, K. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, Dublin Philos. Mag. J. Sci.* **2**, 559–572, <https://doi.org/10.1080/14786440109462720> (1901).
- Cattell, R. B. The scree test for the number of factors. *Multivar. Behav. Res.* **1**, 245, [https://doi.org/10.1207/s15327906mbr0102\\_10](https://doi.org/10.1207/s15327906mbr0102_10) (1966).
- Andersson, H., Baechi, T., Hoehl, M. & Richter, C. Autofluorescence of living cells. *J. microscopy* **191**, 1–7, <https://doi.org/10.1046/j.1365-2818.1998.00347.x> (1998).
- Minamikawa, T. *et al.* Simplified and optimized multispectral imaging for 5-ALA-based fluorescence diagnosis of malignant lesions. *Sci. Reports* **6**, 25530, <https://doi.org/10.1038/srep25530> (2016).
- Warburg, O., Wind, F. & Negelein, E. The metabolism of tumors in the body. *The J. Gen. Physiol.* **8**, 519–530 (1927).
- Van Der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
- Dietel, W., Pottier, R., Pfister, W., Schleier, P. & Zinner, K. 5-Aminolevulinic acid (ALA) induced formation of different fluorescent porphyrins: A study of the biosynthesis of porphyrins by bacteria of the human digestive tract. *J. Photochem. Photobiol. B: Biol.* **86**, 77–86, <https://doi.org/10.1016/j.jphotobiol.2006.07.006> (2007).
- Alston, L., Rousseau, D., Hébert, M., Mahieu-Williams, L. & Montcel, B. Nonlinear relation between concentration and fluorescence emission of protoporphyrin IX in calibrated phantoms. *J. Biomed. Opt.* **23**, 097002, <https://doi.org/10.1117/1.JBO.23.9.097002> (2018).

37. Bellman, R. & Kalaba, R. Reduction of dimensionality, dynamic programming, and control processes. *J. Basic Eng* **83**, 82–84 (1961).
38. Steinbach, M., Ertöz, L. & Kumar, V. The challenges of clustering high dimensional data. In Wille, L. T. (ed.) *New Directions in Statistical Physics: Econophysics, Bioinformatics, and Pattern Recognition*, 273–309, [https://doi.org/10.1007/978-3-662-08968-2\\_16](https://doi.org/10.1007/978-3-662-08968-2_16) (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004).
39. Jolliffe Ian, T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Transactions Royal Soc. A: Math. Phys. Eng. Sci.* **374**, 20150202, <https://doi.org/10.1098/rsta.2015.0202> (2016).
40. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 281–297 (University of California Press, Berkeley, Calif., 1967).
41. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. Ser. B (Methodological)* **39**, 1–22, <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x> (1977).
42. Bishop, C. M. *Pattern Recognition And Machine Learning*, 1st ed. 2006. corr. 2nd printing 2011 edn (Springer-Verlag New York Inc., New York, 2006).

## Acknowledgements

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); Cancéropôle Lyon Auvergne Rhône Alpes (CLARA) within the program “OncoStarter”; Infrastructures d’Avenir en Biologie Santé (ANR-11-INBS-000), within the program “Investissements d’Avenir” operated by the French National Research Agency (ANR). The authors thank the PILoT facility for the support provided on the signal acquisition.

## Author contributions

P.L. conducted the data analysis. L.A., L.M.W., B.M., D.M. and J.G. conducted the experiment. L.A., B.M. and L.M.W. conceived the experiment. C.R., D.R. and B.M. analysed the results. All authors reviewed the manuscript. B.M. and D.R. have equally contributed and jointly supervised the work.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-58299-7>.

**Correspondence** and requests for materials should be addressed to B.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020