

Article

Shapley Value as a Quality Control for Mass Spectra of Human Glioblastoma Tissues

Denis S. Zavorotnyuk ¹, Anatoly A. Sorokin ¹, Stanislav I. Pekov ^{2,3,*}, Denis S. Bormotov ¹, Vasiliy A. Eliferov ¹, Konstantin V. Bocharov ⁴, Eugene N. Nikolaev ^{2,*} and Igor A. Popov ^{1,*}

¹ The Moscow Institute of Physics and Technology, National Research University, 141701 Dolgoprudny, Russia

² Skolkovo Institute of Science and Technology, 121205 Moscow, Russia

³ Siberian State Medical University, 634050 Tomsk, Russia

⁴ V. L. Talrose Institute for Energy Problems of Chemical Physics, N. N. Semenov Federal Research Center for Chemical Physics, Russian Academy of Science, 119334 Moscow, Russia

* Correspondence: stanislav.pekov@forwe.ru (S.I.P.); e.nikolaev@skoltech.ru (E.N.N.); hexapole@gmail.com (I.A.P.)

Abstract: The automatic processing of high-dimensional mass spectrometry data is required for the clinical implementation of ambient ionization molecular profiling methods. However, complex algorithms required for the analysis of peak-rich spectra are sensitive to the quality of the input data. Therefore, an objective and quantitative indicator, insensitive to the conditions of the experiment, is currently in high demand for the automated treatment of mass spectrometric data. In this work, we demonstrate the utility of the Shapley value as an indicator of the quality of the individual mass spectrum in the classification task for human brain tumor tissue discrimination. The Shapley values are calculated on the training set of glioblastoma and nontumor pathological tissues spectra and used as feedback to create a random forest regression model to estimate the contributions for all spectra of each specimen. As a result, it is shown that the implementation of Shapley values significantly accelerates the data analysis of negative mode mass spectrometry data alongside simultaneous improving the regression models' accuracy.

Keywords: ambient ionization mass spectrometry; Shapley value; classification



Citation: Zavorotnyuk, D.S.; Sorokin, A.A.; Pekov, S.I.; Bormotov, D.S.; Eliferov, V.A.; Bocharov, K.V.; Nikolaev, E.N.; Popov, I.A. Shapley Value as a Quality Control for Mass Spectra of Human Glioblastoma Tissues. *Data* **2023**, *8*, 21. <https://doi.org/10.3390/data8010021>

Academic Editors: Muhammad Irfan, Thompson Sarkodie-Gyan and Wahyu Caesarendra

Received: 1 November 2022

Revised: 24 December 2022

Accepted: 5 January 2023

Published: 16 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The explosion in interest surrounding ambient ionization mass spectrometry has turned fingerprinting into a prominent method for a variety of clinical implementations [1,2]. The absence of time-consuming sample preparations and separation steps has allowed for it to be possible to integrate ambient ionization mass spectrometry into routine diagnostic and surgical pipelines [3,4]. The intraoperative differentiation and evaluation of resected tissues is becoming more and more in demand in oncological surgery as a helpful decision-making technique, as well as a way to accelerate biopsy examinations [4–6]. Unlike widespread approaches based on single biomarker identification [7–9], molecular profiling requires the detection of a complex molecular signature of cancer tissue [10–12]. This means that an analysis of peak-rich mass spectra is required to determine distinctive features, regardless of the exact ion composition in the detected peaks. However, such an analysis inevitably leads to high data dimensionality. In such cases, it is common to use averages over certain axes, which reduces the input dataset and masks possible inconsistencies therein.

The automatic processing of high-dimensional data is required for the clinical implementation of the proposed molecular profiling techniques; thus, more complex data analysis algorithms are becoming more crucial [13–15]. At the same time, the more complex the algorithm used, the more samples should be subjected to a mass spectrometric analysis. However, the typical number of available samples [16,17] (i.e., individual patients

or biopsy specimens, which usually count in the dozens to low hundreds) is lower than the number of characteristics describing the model (i.e., peaks and scans, which could count in the hundreds or thousands) [18,19]. On the other hand, complex algorithms are sensitive to the quality of the input data, and a simple increase in the number of spectra from each individual specimen could increase the percentage of inadequate data caused by experimental instabilities or sample exhaustion [20].

The quality of individual spectrum scans can be determined with the magnitude of the total ion current, the number of peaks, or with the signal-to-noise ratio. Despite the large number of tools designed for this purpose, the results should be manually inspected by an expert [21–23], since, for instance, the polarity mode, the resolution of the detector, and the scanning range can cause large differences in the parameters used for the control of spectra quality. Therefore, a more universal and quantitative indicator, which does not depend as strongly on the experimental conditions, is required for the automated processing and analysis of mass spectrometric data. To obtain such an indicator, one of the quantitative data valuation methods can be used. The set of these methods consists of influence functions [24], leave-one-out validation [25], reinforcement learning, and Shapley's value computation [12,26].

In this work, the Shapley value as an indicator of the quality of the individual scan (mass spectrum) is proposed. In supervised machine learning, given the training dataset, the learning algorithm, and the means of assessing the learning algorithm performance, we could obtain the importance of one particular data point as a measure of whether excluding this data point from the training dataset decreases or increases the performance of the learning algorithm. At the same time, this measure should satisfy the three axioms: symmetry, efficiency, and "the aggregation law" [27]. It was previously stated and proven [26] that the method which satisfies the axioms listed above must have the form

$$\phi_i = C \sum_{S \subseteq D - \{i\}} \frac{V(S \cup \{i\}) - V(S)}{\binom{n-1}{|S|}} \quad (1)$$

where C is an arbitrary constant and the sum is computed over all subsets of the dataset D with the i -th data point excluded. The value ϕ_i is called the Shapley value of the i -th data point. In this assay, the Shapley value is essentially the contribution of the i -th scan into a metric that characterizes the quality of the model built using that scan among others in a whole dataset. The classification model is fitted on previously collected human brain tumor mass spectra [17] and is based on the hypothesis of there being an alteration in the lipid metabolism during the malignant transformation of glial cells [28,29]. The Shapley value calculation consists of determining how the accuracy (or any other metric) of the machine learning model changes if this particular scan is removed from the training set.

2. Materials and Methods

2.1. Experimental Data

Tissue samples were provided by the N.N. Burdenko NSPCN and analyzed under a protocol approved by the N.N. Burdenko NSPCN Institutional Review Board. Brain tumor tissues ($n = 208$) were resected during elective surgeries of glioblastoma patients. Nontumor pathological tissues ($n = 40$) were resected in the course of the surgical treatment of drug-resistant epilepsy. A signed informed consent form, filled out in accordance with the requirements of the local ethical committee, specifically noting that all removed tissues could be used for further research, was obtained from all patients before surgery. The study was conducted in accordance with the Helsinki Declaration, as revised in 2013. All procedures were carried out according to the relevant guidelines and regulations.

All dissected tissue was anonymized, examined by a professional pathologist, and placed in normal saline, frozen, and stored at -80 °C until analysis. The samples were analyzed using an inline cartridge extraction (ICE) ambient ionization mass spectrometry approach [17]. Briefly, a freshly thawed tissue sample was cut into an approximately

1 mm³ large sample and placed into disposable stainless-steel cartridges. A high voltage (3.5 ± 1 kV, tuned to form a stable single-jet Taylor cone) and solvent flow (3 μ L/min) were then applied through the cartridge to obtain a stable ion current. In total, 90% HPLC-grade methanol supplemented with 0.1% acetic acid was used as an extraction solvent. The solvents and acetic acid were obtained from Merck (Merck KGaA, Darmstadt, Germany). The acquisition of mass spectra was performed on the Thermo LTQ XL Orbitrap ETD mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA). Samples were analyzed in the negative and positive modes in the m/z ranges 500–1000 m/z .

2.2. Shapley Data

The mass spectra were preliminarily aligned to the spectrum with the maximum total ion current. The alignment was performed within each class of spectra (glioblastoma/nontumor pathology) separately. After the alignment, a total matrix of peak intensities was obtained with a size of $13,611 \times 198$ for the spectra of negative ions and $15,102 \times 200$ for the spectra of positive ions. The calculation procedure was carried out for each set of spectra separately.

The Shapley values were calculated for 1200 scans, which were selected randomly from the total peak intensity matrices. Scan contributions were evaluated on validation sets of 1500 scans.

The sets were compiled in such a way that scans from both groups of spectra were equally included in each set. It was also ensured that scans belonging to the same specimen were not included in the training and verification sets at the same time.

The calculation process was as follows: a random rearrangement was performed in the original training set of the scans. For a newly ordered set, the learning algorithm determined the contribution of each scan to a given performance metric. The traditional logistic regression model was used as a training algorithm, and the accuracy of the class prediction for a validating set of scans was calculated as a quality metric for the model. The contribution of a scan is defined as the change in prediction accuracy when adding this scan to the training set of predictors compared to the previous accuracy value, averaged over the number of scans before the addition. The procedure for calculating the permutations was repeated until the sum of the relative changes in the contributions of the scans over the last 100 permutations became less than a certain threshold (ShapTolerance), which was fixed at a value of 0.5 in this study. The scan contributions obtained for the last permutation were taken as the Shapley values. Data on the histological diagnoses of patients were used as a response in creating the model.

Determining Shapley values can take a long time, even for a set as small as 1200 scans, so the permutation calculation procedure was performed in parallel on 8 CPU cores, and the ShapTolerance threshold was checked once for every thousand permutations.

To extend the Shapley analysis to the whole dataset, the calculated Shapley values were used as feedback to create a regression model to estimate the contributions for all scans. The model was built using the random forest method with cross-validation (5 splits with 3 repetitions).

To determine the influence of the scanned Shapley values on the accuracy of the resulting model, 1200 scans were selected from the entire set for the calculated Shapley values and 4000 scans for the predicted values. The selection of scans was performed so that each set contained an approximately equal number of glioblastoma scans and nontumor samples. Selected scans were aligned in ascending order, according to the Shapley value, and removed from the training set according to that order. The remaining scans were used to create a cross-validated logistic regression model. After that, for the resulting model, the accuracy of predicting classes from the training set of scans was estimated.

Shapley values and their processing were calculated in the R environment versions 3.4.4 and 4.0.4 using the R packages MALDIquant, glmnet [30], doParallel [31], caret [32], and ggplot2 [33]. The negative ion spectra were processed on a 12-core desktop computer with 32 GB RAM running Ubuntu 16.04 OS, and the positive ion spectra were processed on

a 16-core desktop computer with 32 GB RAM running Ubuntu 20.04 OS. The Shapley data calculation was performed simultaneously on 8 CPU cores.

3. Results

3.1. Calculated Shapley Values

The iterative process of calculating the Shapley values was limited to a ShapTolerance of 0.5. As a result, Shapley values of 1200 scans were obtained for two sets of spectra obtained in the negative and positive ion modes, with their model performance estimated on a validation set consisting of 1500 scans. The distribution of calculated values is shown in Figure 1.

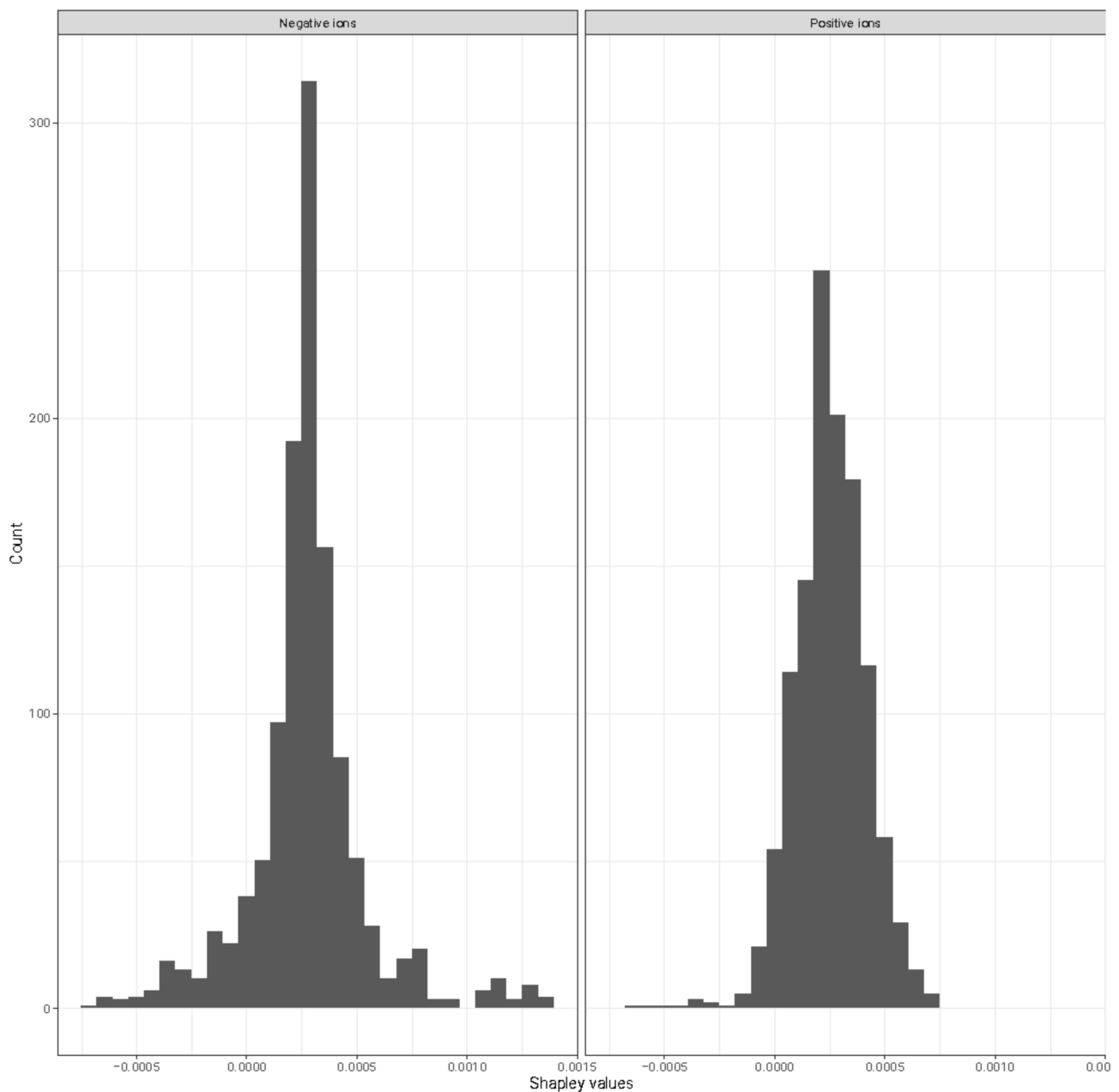


Figure 1. Distribution of calculated Shapley values. The values obtained for the negative and positive ion modes are shown on the left and right hand sides, respectively.

Excluding scans with negative Shapley values as predictors affected the accuracy of the classification model, as shown in Figure 2. It can be seen that, when scans with negative

Shapley values were excluded (the area to the left of the vertical red line), the accuracy of the model increased.

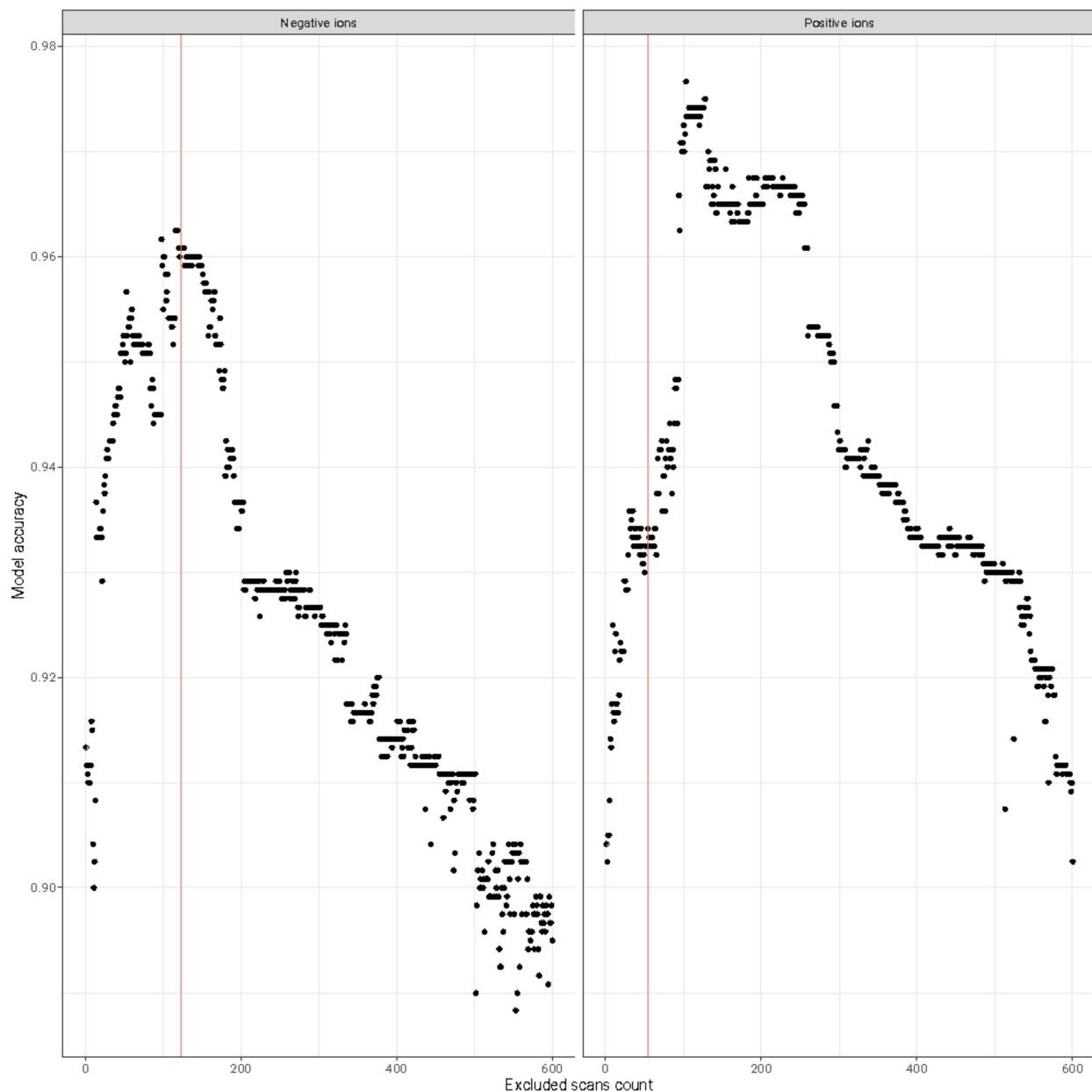


Figure 2. Change in classification accuracy in the scan exclusion procedure. The plots for the negative and positive ion regime spectra are shown on the left and right hand sides, respectively. The red vertical line shows the border separating the areas of negative (left) and positive (right) Shapley values.

3.2. Shapley Value Modeling

The time required to calculate the Shapley values can be measured in days and weeks for a large set of scans; therefore, in this study, values were calculated for a smaller set of scans and the random forest regression model was trained to predict the Shapley values for the entire set of available experimental data. The parameters of the obtained models are presented in Table 1. There was a good agreement between the predicted and actual Shapley values shown in Figure 3, although it should be noted that the deviation was bigger in the area of the highest Shapley values, which was less important in terms of quality control.

Table 1. Parameters of regression models. Neg and Pos are models for the Shapley data for the spectra of negative and positive ions, respectively.

Scan Set	Number of Predictors	RMSE	R2	MAE
Neg	198	8.473×10^{-5}	0.8972	5.0221×10^{-5}
Pos	200	6.550×10^{-5}	0.8411	4.6361×10^{-5}

Using the models, the Shapley values were calculated for the entire set of scans. The distribution of the predicted Shapley data was generally the same as the distribution of the calculated ones (see Supplementary Figures S1 and S2).

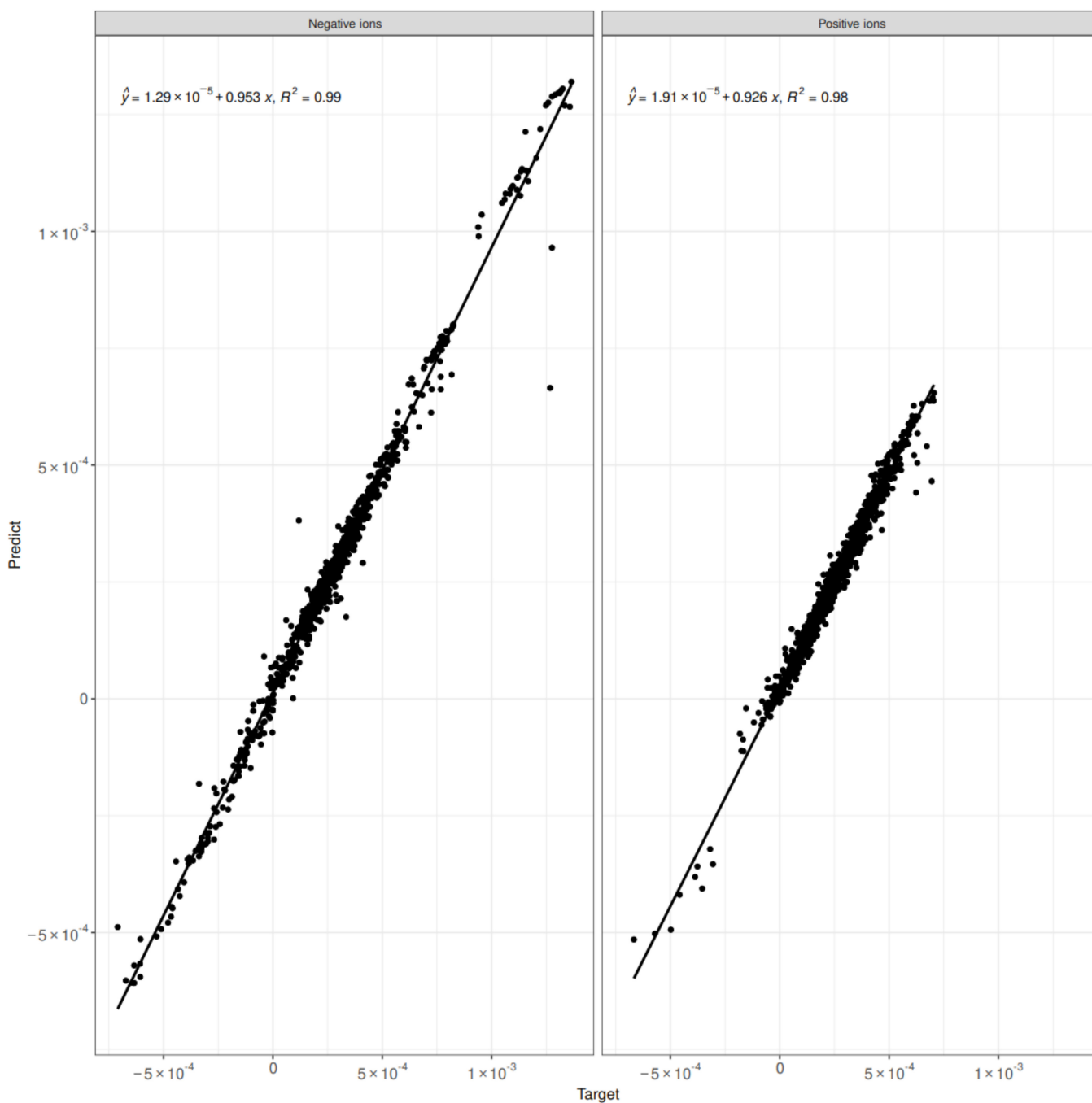


Figure 3. Correspondence of the predicted Shapley values with the calculated ones. The model for the mass spectrometric scan data in the negative (left) and positive (right) ion modes.

The Shapley values obtained using regression models were used to exclude from the analysis mass spectrometric scans, which corresponded to negative and zero Shapley values. To evaluate the results, the construction time and accuracy of the logistic regression model were measured for two cases: the entire set of scans and the set of only those scans that had positive Shapley values. The data presented in Table 2 show a decrease in calculation time and an increase in prediction accuracy when using only the scans with positive Shapley values.

Table 2. Comparison of construction time and accuracy of models when only the scans with positive Shapley values were included in the analysis.

Dataset	Duration (Seconds)		Model Accuracy	
	Negative Mode	Positive Mode	Negative Mode	Positive Mode
General	38.738	24.070	0.9626	0.9860
Shapley-filtered	15.730	23.354	0.9719	0.9881

4. Discussion

The Shapley values could be used to evaluate the relevance of the respective scans for further consideration. By definition, adding scans with negative Shapley values worsened the quality of the resulting model, positive values improved it, and zero values did not affect the simulation results. Therefore, scans with negative and zero Shapley values should be excluded from further consideration.

Figure 2 shows that the accuracy of the class prediction increased when the predictors with low Shapley values were excluded until a certain maximum value was reached, after which the accuracy started to decrease due to a decrease in the size of the training dataset.

The results of this study showed that the Shapley values calculated using a regression model for the entire set of scans could also be used as a reliable scan quality metric.

For the spectra of negative ions, the number of scans with negative Shapley values was 10% of the calculated and 16% of the predicted values; the same data for positive ions were 4% for the calculated values and 3% for the predicted ones. This was in great accordance with previous results indicating that positive spectra were much more stable compared to negative ones [20]. This could also explain that the Shapley values for the positive ion mode scans were smaller in the absolute value than for the negative ion mode.

The obtained results indicated that the Shapley value, as proposed in Ref. [26], could be used as a quality metric for mass spectrum scans and that the regression model built on a small subset of data could provide a good estimation of Shapley values to rank scans according to their quality. The application of the Shapley value quality filter drastically reduced the model training time and increased the classification accuracy.

Nevertheless, Data Shapley should not be treated as the outlier detection algorithm, as they did not use assumptions about the data distribution and the data normality. The scan with the negative Shapley value could either be an outlier, for instance, in the sense of the total ion current, or the scan belonging to the part of the sample tissue corresponding to the tumor boundary and, thus, containing malignant tumor cells and healthy brain cells. Supplementary Figures S3 and S4 with the PCA diagrams of the negative ion scans colored with the positive and nonpositive Shapley values showed that nonpositive Shapley value scans could hardly be treated as outliers.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/data8010021/s1>, Figures S1 and S2: plots of change in prediction accuracy with the exclusion of samples from the predicted Shapley values for different polarities; Figures S3 and S4: PCA diagrams for the scans of the negative ions colored with the positive and non-positive Shapley values.

Author Contributions: Conceptualization, A.A.S.; methodology, D.S.Z. and A.A.S.; software, D.S.Z.; validation, D.S.Z. and A.A.S.; formal analysis, D.S.Z.; investigation, S.I.P., D.S.B. and V.A.E.; resources, E.N.N. and I.A.P.; data curation, D.S.Z.; writing—original draft preparation, D.S.Z.; writing—review and editing, A.A.S., S.I.P. and D.S.B.; visualization, D.S.Z.; supervision, A.A.S. and I.A.P.; project administration, K.V.B.; funding acquisition, I.A.P. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by the Ministry of Science and Higher Education of the Russian Federation, project no. 0714-2020-0006, agreement no. 075-00337-20-02. The research used the equipment of the Shared Research Facilities of the Semenov Federal Research Center for Chemical Physics RAS.

Institutional Review Board Statement: The study was approved by the N.N. Burdenko NSPCN and analyzed under a protocol approved by the N.N. Burdenko NSPCN Institutional Review Board (order 40 from 12 April 2016, revised with order 131 from 17 July 2018). A signed informed consent form, filled out in accordance with the requirements of the local ethical committee, specifically noting that all removed tissues could be used for further research, was obtained from all patients before surgery. The study was conducted in accordance with the Helsinki Declaration, as revised in 2013.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data described in the manuscript are unavailable, in accordance with the indication of the Ethics Committee. The partial dataset is available through ref. [16]. The source code of the module used to obtain the calculated Shapley values is accessible on the GitHub repository via the link <https://github.com/lptolik/ShapleyDataR> (accessed on 10 July 2022).

Acknowledgments: The authors would like to thank Ekaterina A. Bormotova and Mariya M. Derkach for their valuable help in the preparation of the manuscript.

Conflicts of Interest: The authors declare no competing interests.

References

1. Li, L.-H.; Hsieh, H.-Y.; Hsu, C.-C. Clinical Application of Ambient Ionization Mass Spectrometry. *Mass Spectrom.* **2017**, *6*, S0060. [[CrossRef](#)] [[PubMed](#)]
2. Rankin-Turner, S.; Reynolds, J.C.; Turner, M.A.; Heaney, L.M. Applications of Ambient Ionization Mass Spectrometry in 2021: An Annual Review. *Anal. Sci. Adv.* **2022**, *3*, 67–89. [[CrossRef](#)]
3. Pekov, S.I.; Zhvansky, E.S.; Eliferov, V.A.; Sorokin, A.A.; Ivanov, D.G.; Nikolaev, E.N.; Popov, I.A. Determination of Brain Tissue Samples Storage Conditions for Reproducible Intraoperative Lipid Profiling. *Molecules* **2022**, *27*, 2587. [[CrossRef](#)] [[PubMed](#)]
4. Pekov, S.I.; Bormotov, D.S.; Nikitin, P.V.; Sorokin, A.A.; Shurkhay, V.A.; Eliferov, V.A.; Zavorotnyuk, D.S.; Potapov, A.A.; Nikolaev, E.N.; Popov, I.A. Rapid Estimation of Tumor Cell Percentage in Brain Tissue Biopsy Samples Using Inline Cartridge Extraction Mass Spectrometry. *Anal. Bioanal. Chem.* **2021**, *413*, 2913–2922. [[CrossRef](#)] [[PubMed](#)]
5. Iwano, T.; Yoshimura, K.; Inoue, S.; Odate, T.; Ogata, K.; Funatsu, S.; Tanihata, H.; Kondo, T.; Ichikawa, D.; Takeda, S. Breast Cancer Diagnosis Based on Lipid Profiling by Probe Electrospray Ionization Mass Spectrometry. *Br. J. Surg.* **2020**, *107*, 632–635. [[CrossRef](#)]
6. Giordano, S.; Siciliano, A.M.; Donadon, M.; Soldani, C.; Franceschini, B.; Lleo, A.; di Tommaso, L.; Cimino, M.; Torzilli, G.; Saiki, H.; et al. Versatile Mass Spectrometry-Based Intraoperative Diagnosis of Liver Tumor in a Multiethnic Cohort. *Appl. Sci.* **2022**, *12*, 4244. [[CrossRef](#)]
7. Pirro, V.; Llor, R.S.; Jarmusch, A.K.; Alfaro, C.M.; Cohen-Gadol, A.A.; Hattab, E.M.; Cooks, R.G. Analysis of Human Gliomas by Swab Touch Spray-Mass Spectrometry: Applications to Intraoperative Assessment of Surgical Margins and Presence of Oncometabolites. *Analyst* **2017**, *142*, 4058–4066. [[CrossRef](#)]
8. Shamraeva, M.A.; Bormotov, D.S.; Shamarina, E.V.; Bocharov, K.V.; Peregudova, O.V.; Pekov, S.I.; Nikolaev, E.N.; Popov, I.A. Spherical Sampler Probes Enhance the Robustness of Ambient Ionization Mass Spectrometry for Rapid Drugs Screening. *Molecules* **2022**, *27*, 945. [[CrossRef](#)]
9. Del Mar Boronat Ena, M.; Cowan, D.A.; Abbate, V. Ambient Ionization Mass Spectrometry Applied to New Psychoactive Substance Analysis. *Mass. Spectrom. Rev.* **2021**, *42*, 3–34. [[CrossRef](#)]
10. Ogrinc, N.; Attencourt, C.; Colin, E.; Boudahi, A.; Tebbakha, R.; Salzet, M.; Testelin, S.; Dakpé, S.; Fournier, I. Mass Spectrometry-Based Differentiation of Oral Tongue Squamous Cell Carcinoma and Nontumor Regions With the SpiderMass Technology. *Front. Oral Health* **2022**, *3*, 827360. [[CrossRef](#)]
11. King, M.E.; Zhang, J.; Lin, J.Q.; Garza, K.Y.; DeHoog, R.J.; Feider, C.L.; Bensussan, A.; Sans, M.; Krieger, A.; Badal, S.; et al. Rapid Diagnosis and Tumor Margin Assessment during Pancreatic Cancer Surgery with the MasSpec Pen Technology. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2104411118. [[CrossRef](#)]

12. Xie, Y.R.; Castro, D.C.; Bell, S.E.; Rubakhin, S.S.; Sweedler, J.V. Single-Cell Classification Using Mass Spectrometry through Interpretable Machine Learning. *Anal. Chem.* **2020**, *92*, 9338–9347. [[CrossRef](#)]
13. Boiko, D.A.; Kozlov, K.S.; Burykina, J.V.; Ilyushenkova, V.V.; Ananikov, V.P. Fully Automated Unconstrained Analysis of High-Resolution Mass Spectrometry Data with Machine Learning. *J. Am. Chem. Soc.* **2022**, *144*, 14590–14606. [[CrossRef](#)]
14. Piras, C.; Hale, O.J.; Reynolds, C.K.; Jones, A.K.B.; Taylor, N.; Morris, M.; Cramer, R. LAP-MALDI MS Coupled with Machine Learning: An Ambient Mass Spectrometry Approach for High-Throughput Diagnostics. *Chem. Sci.* **2022**, *13*, 1746–1758. [[CrossRef](#)]
15. Liebal, U.W.; Phan, A.N.T.; Sudhakar, M.; Raman, K.; Blank, L.M. Machine Learning Applications for Mass Spectrometry-Based Metabolomics. *Metabolites* **2020**, *10*, 243. [[CrossRef](#)]
16. Zavorotnyuk, D.S.; Pekov, S.I.; Sorokin, A.A.; Bormotov, D.S.; Levin, N.; Zhvansky, E.; Semenov, S.; Strelnikova, P.; Bocharov, K.V.; Vorobiev, A.; et al. Lipid Profiles of Human Brain Tumors Obtained by High-Resolution Negative Mode Ambient Mass Spectrometry. *Data* **2021**, *6*, 132. [[CrossRef](#)]
17. Pekov, S.I.; Eliferov, V.A.; Sorokin, A.A.; Shurkhay, V.A.; Zhvansky, E.S.; Vorobyev, A.S.; Potapov, A.A.; Nikolaev, E.N.; Popov, I.A.; Alexander, S.V.; et al. Inline Cartridge Extraction for Rapid Brain Tumor Tissue Identification by Molecular Profiling. *Sci. Rep.* **2019**, *9*, 18960. [[CrossRef](#)]
18. Thomas, S.A.; Race, A.M.; Steven, R.T.; Gilmore, I.S.; Bunch, J. Dimensionality Reduction of Mass Spectrometry Imaging Data Using Autoencoders. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*; IEEE: New York, NY, USA, 2016. [[CrossRef](#)]
19. Zhvansky, E.; Sorokin, A.; Shurkhay, V.; Zavorotnyuk, D.; Bormotov, D.; Pekov, S.; Potapov, A.; Nikolaev, E.; Popov, I. Comparison of Dimensionality Reduction Methods in Mass Spectra of Astrocytoma and Glioblastoma Tissues. *Mass Spectrom.* **2021**, *10*, A0094. [[CrossRef](#)]
20. Zhvansky, E.S.; Eliferov, V.A.; Sorokin, A.A.; Shurkhay, V.A.; Pekov, S.I.; Bormotov, D.S.; Ivanov, D.G.; Zavorotnyuk, D.S.; Bocharov, K.V.; Khaliullin, I.G.; et al. Assessment of Variation of Inline Cartridge Extraction Mass Spectra. *J. Mass Spectrom.* **2021**, *56*, e4640. [[CrossRef](#)]
21. Zhvansky, E.S.; Pekov, S.I.; Sorokin, A.A.; Shurkhay, V.A.; Eliferov, V.A.; Potapov, A.A.; Nikolaev, E.N.; Popov, I.A. Metrics for Evaluating the Stability and Reproducibility of Mass Spectra. *Sci. Rep.* **2019**, *9*, 914. [[CrossRef](#)]
22. Gibb, S.; Strimmer, K. MALDIquant: A Versatile R Package for the Analysis of Mass Spectrometry Data. *Bioinformatics* **2012**, *28*, 2270–2271. [[CrossRef](#)] [[PubMed](#)]
23. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinform.* **2010**, *11*, 395. [[CrossRef](#)] [[PubMed](#)]
24. Koh, P.W.; Liang, P. Understanding Black-Box Predictions via Influence Functions. In Proceedings of the 34th International Conference on Machine Learning, Sydney, NSW, Australia, 6–11 August 2017; pp. 1885–1894.
25. Molinaro, A.M.; Simon, R.; Pfeiffer, R.M. Prediction Error Estimation: A Comparison of Resampling Methods. *Bioinformatics* **2005**, *21*, 3301–3307. [[CrossRef](#)] [[PubMed](#)]
26. Ghorbani, A.; Zou, J. Data Shapley: Equitable Valuation of Data for Machine Learning. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; Volume 97, pp. 2242–2251.
27. Shapley, L.S. A value for n-person games. *Contrib. Theory Games* **1953**, *2*, 307–317.
28. Sorokin, A.; Shurkhay, V.; Pekov, S.; Zhvansky, E.; Ivanov, D.; Kulikov, E.E.; Popov, I.; Potapov, A.; Nikolaev, E. Untangling the Metabolic Reprogramming in Brain Cancer: Discovering Key Molecular Players Using Mass Spectrometry. *Curr. Top. Med. Chem.* **2019**, *19*, 1521–1534. [[CrossRef](#)]
29. Pekov, S.I.; Sorokin, A.A.; Kuzin, A.A.; Bocharov, K.V.; Bormotov, D.S.; Shivalin, A.S.; Shurkhay, V.A.; Potapov, A.A.; Nikolaev, E.N.; Popov, I.A. Analysis of Phosphatidylcholines Alterations in Human Glioblastomas Ex Vivo. *Biochem. Moscow Suppl. Ser. B Biomed. Chem.* **2021**, *15*, 241–247. [[CrossRef](#)]
30. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)]
31. Microsoft Corporation and Steve Weston. doParallel: Foreach Parallel Adaptor for the ‘parallel’ Package. R package version 1.0.17. 2022. Available online: <https://CRAN.R-project.org/package=doParallel> (accessed on 10 July 2022).
32. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
33. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; ISBN 978-3-319-24277-4.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.