BioMedInformatics

MDPI

Article

State-of-the-Art Explainability Methods with Focus on Visual Analytics Showcased by Glioma Classification

Milot Gashi [1,†], Matej Vuković [1,†], Nikolina Jekic [2,†], Stefan Thalmann [3], Andreas Holzinger [4,5,6], Claire Jean-Quartier [4,5,*] and Fleur Jeanquartier [4,*]

1 Pro2Future, Inffeldgasse 25F, 8010 Graz, Austria; milot.gashi@pro2future.at (M.G.); matej.vukovic@pro2future.at (M.V.)
2 Institute of Computer Graphics and Knowledge Visualisation, Graz University of Technology, 8010 Graz, Austria; nikolinajekic@hotmail.com
3 Business Analytics and Data Science Center, University of Graz, 8010 Graz, Austria; stefan.thalmann@uni-graz.at
4 Human-Centered AI Lab (Holzinger Group), Institute for Medical Informatics, Statistics and Documentation, Medical University Graz, 8036 Graz, Austria; andreas.holzinger@human-centered.ai
5 Institute for Data Science and Interactive Systems, Graz University of Technology, 8010 Graz, Austria
6 xAI Lab, Alberta Machine Intelligence Institute, University of Alberta, Edmonton, AB T6G 2E8, Canada
* Correspondence: c.jeanquartier@hci-kdd.org (C.J.-Q.); f.jeanquartier@hci-kdd.org (F.J.)
† These authors contributed equally to this work.

Abstract: This study aims to reflect on a list of libraries providing decision support to AI models. The goal is to assist in finding suitable libraries that support visual explainability and interpretability of the output of their AI model. Especially in sensitive application areas, such as medicine, this is crucial for understanding the decision-making process and for a safe application. Therefore, we use a glioma classification model's reasoning as an underlying case. We present a comparison of 11 identified Python libraries that provide an addition to the better known SHAP and LIME libraries for visualizing explainability. The libraries are selected based on certain attributes, such as being implemented in Python, supporting visual analysis, thorough documentation, and active maintenance. We showcase and compare four libraries for global interpretations (ELI5, Dalex, InterpretML, and SHAP) and three libraries for local interpretations (Lime, Dalex, and InterpretML). As use case, we process a combination of openly available data sets on glioma for the task of studying feature importance when classifying the grade II, III, and IV brain tumor subtypes glioblastoma multiforme (GBM), anaplastic astrocytoma (AASTR), and oligodendroglioma (ODG), out of 1276 samples and 252 attributes. The exemplified model confirms known variations and studying local explainability contributes to revealing less known variations as putative biomarkers. The full comparison spreadsheet and implementation examples can be found in the appendix.

Keywords: explainable artificial intelligence; visualisation; SHAP; feature importance; Python; glioma

Citation: Gashi, M.; Vukovic, M.; Jekic, N.; Thalmann, S.; Holzinger, A.; Jean-Quartier, C.; Jeanquartier, F. State-of-the-Art Explainability Methods with Focus on Visual Analytics Showcased by Glioma Classification. *Biomedinformatics* **2022**, 2, 139–158. https://doi.org/10.3390/biomedinformatics2010009

Academic Editors: Jörn Lötsch and Alfred Ultsch

Received: 30 December 2021
Accepted: 13 January 2022
Published: 19 January 2022

1. Introduction

In recent years, extensive benefits to different application areas have been offered due to successfully applying machine learning (ML) algorithms. In particular, the success of deep learning (DL) approaches are transforming the way we approach real-world tasks performed by humans. ML and DL establish artificial intelligence (AI) models which can be applied in many different fields of research such as healthcare [1], cancer classification [2–4], autonomous robots and vehicles [5], image processing [6], manufacturing, and many more [7–10], thus enhancing and providing various benefits in the corresponding fields. Moreover, these models resulting from ML are suitable for performing different tasks, such as recommendation, ranking, forecasting, classification, or clustering. The variety and the nature of these approaches make them complex to understand and interpret. In the

literature, AI models are generally known as black-box, particularly if they result from ML or DL [11]. The opaqueness of such models has negative effects on user acceptance [12]. It also limits the application in sensitive cases such as medicine, finance, or law, where explanations are crucial for users to understand and interpret results in order to effectively manage and use the underlying algorithms [11,13]. From the legal perspective, most applications of AI in medicine are defined as high-risk use cases of AI according to the legal framework for regulating the use of AI proposed by the European Commission (European Commission, 2021). In case of a high-risk application it is required to provide transparency and clear and comprehensible information about the system and its decisions to the user. Such explanations are also dictated by the European General Data Protection Regulation, but also by Californian law (Title 1.81.8. Automated Decision Systems Accountability Act of 2020). Traditionally, software validation or IT auditing is applied in order to fulfill the legal and, in many cases, compliance requirements. However, due to the black-box characteristic resulting from ML and DL, traditional approaches are no longer sufficient, and new guidelines and approaches are needed [14]. In this regard, Explainable Artificial Intelligence (xAI) is proposed as a technical solution, and the first successful validations are already performed in sensitive areas such as pharmaceutical production [15]. In addition, xAI can also increase user acceptance, and the application rate of these models [12]. Thus, xAI approaches seem promising to handle this challenge from a technical perspective.

xAI approaches aim to extract knowledge of what the AI algorithm learned during training and how the decision for particular or new instances are generated during the prediction process. xAI mainly focuses on two methods to provide an explanation at a different level of detail: local and global explainability. Local explainability aims to explain particular prediction output, e.g., prediction of single instances. We find many different techniques focusing on local explainability in the xAI literature [16,17]. On the other hand, the goal of global explainability is to explain the overall model behavior, rather than a particular instance. Global methods are extensively applied in different domains, such as health care [18,19], manufacturing [20], administration of justice [21], or biomedical science [22]. These methods mainly rely on dimension reduction and visualization techniques to provide an intuitive explanation to humans. Visualizing a process helps us understand ML models and decision-making processes in a more intuitive way [23]. Moreover, visual inspection is considered as an easy and fast way to recognize new knowledge while analyzing complex processes [24]. As a result, visualization in the context of xAI is widely applied, thus facilitating the interpretation process of black-box models [11,25,26]. Users benefit from visual analytic (VA) systems for xAI [27]. Many of these methods are implemented in Python or R and are openly available [17,28,29]. This helps researchers and, in general, the data-driven community to use and enhance further state-of-the-art solutions. Some existing methods have already been summarized [30–32]. However, a comparison of ease of use regarding implementation, as well as details on visualization features, is missing.

In this paper, we report on a structured review to investigate the state of the art of mature xAI libraries incorporating VA features. We analyzed the characteristics of xAI libraries with respect to ease of installation and documentation. The comparison is use-case driven: we compare and rank selected libraries regarding their VA capabilities for global and local explainability in general. In particular, we explore different implementations of lime and SHAP approaches and apply selected libraries for the use case of investigating glioma classification based on several clinical and genetic variables. We thereby showcase the applicability of xAI on and supporting the biomedical knowledge creation process.

1.1. Classification of Diffuse Glioma

Classification of glioma subtypes is important for therapy decisions and is based on gene variations [33]. This list of central nervous system tumors has been introduced by the World Health Organization and has been updated recently [34]. The community-driven cancer classification platform Oncotree has been developed as clinical decision support system for oncology research and precision medicine and allows for dynamic granular-

ity [35]. For example, grading of diffuse gliomas (DIFG) is still an ongoing discussion and momentarily defined by tumor nomenclature [36]. The process involves molecular and histological features in order to revise risk stratification. Common molecular biomarkers used for clinical classification of glioma include α-thalassemia/mental retardation syndrome X-linked (ATRX), isocitrate dehydrogenase 1 (IDH1), tumor protein p53 (TP53), telomerase reverse transcriptase (TERT), and phosphatase and tensin homolog (PTEN) or the epidermal growth factor receptor (EGFR) among others [34,37]. We have recently highlighted age-based differences in brain tumor diseases using an explainable classification approach [22]. We now extend our studies to include several xAI methods for classifying DIFGs.

1.2. Theoretical Background on xAI

xAI is defined for the first time in 2004 by Can Lent et al. [38] as a research field that explains the behavior AI models in a more understandable way. However, focus on the topic of xAI has been recently increasing [32] due to increased attention and improvements around the topic of AI/ML across different fields. However, along with the high accuracy results, a more human-centric explanation of the decision-making process of these models is required. This leads the focus toward xAI in the current age. Furthermore, the increase in complexity of ML models has lead to the requirement for developing algorithmic decision-making such as fairness, accountability, and transparency (FAT) principles [39] which are especially evident in highly regulated and mission-critical scenarios.

There are several perspectives on the explainability of an AI model (e.g., scope, stage, problem type, etc.). The scope perspective regards the global and local view on model explanations. AI models can be explained either at the global level or local level. Global level interpretation is known as global interpretability in the literature [32], where the entire model behavior is analyzed e.g., feature importance. Global level interpretability summarizes the impact of input features on the model, as well as the model as a whole, while the local interpretation is defined as local interpretability, and it aims to understand the behavior of single predictions and decisions made by the model.

Another perspective on the explainability of an AI model is associated with the type of AI model itself. Overall, two types of models exist, white-box and black-box models. White-box models are made to be explainable by design, resulting in no requirement of additional xAI methods for the model to be explainable. Contrarily, black-box models are not explainable by design, so other techniques have to be applied to extract reasoning for certain decisions and predictions.

In regard to xAI methods, a recent study [32] reviewed more than 200 scientific articles that aimed to develop new methods for explainability. However, discussing these methods and other xAI concepts falls outside of the scope of this paper. We encourage the reader to consult the work discussed in [30–32] for more details about these concepts.

2. Materials and Methods

2.1. Dataset

Data on glioma samples were downloaded from cbioportal [40,41] with filtering the 6 studies gbm_mayo_pdx_sarkaria_2019, gbm_tcga_pub2013, glioma_mskcc_2019, lgg_tcga, lgg_ucsf_2014, and odg_msk_2017. Only data with the 7 attributes "Oncotree Code", "Mutation Count", "Overall Survival (Months)", "Overall Survival Status", "Sex", "Somatic Status", and "Diagnosis Age" were used. Sample rows without complete data have been removed. Data were extended with gene mutation data of the top 246 mutated genes within selected studies.

The top three diffuse glioma (DIFG) subtypes (Glioblastoma multiforme (GBM), Anaplastic Astrocytoma (AASTR), and Oligodendroglioma (ODG)) were further selected and analyzed within this work. We filtered and further processed data for model building comprising of 1276 sample rows with 253 columns out of the 5 studies gbm_mayo_pdx_sarkaria_2019, gbm_tcga_pub2013, glioma_mskcc_2019, lgg_tcga, and lgg_ucsf_2014. The Oncotree Code

was selected as the target and the other 252 data columns were selected as features, with 872 GBM sample rows, 234 AASTR sample rows, and 170 ODG sample rows. The data preprocessing and model building can be found on https://github.com/mathabaws/SOTA_xAI_Visual_analytics/blob/main/notebooks/diffuseglioma-dataset-processing.ipynb (accessed on 12 January 2022).

2.2. Implementation

We conducted a structured review with the goal of investigating current developments and the state of the art xAI libraries focusing on model interpretation and visualization techniques. State of the art means most up to date, publicly available, implemented consistently with the requirement of current software technology, and following common Python patterns. Moreover, this review aims to investigate various relevant aspects of xAI libraries such as maturity level, documentation, supported programming languages, models and different machine learning tasks, support for data types, etc. The structured review closely follows the methodology for Structured Literature Review (SLR) from Webster and Watson [42]. Additionally, we take necessary attributes for a software selection process into account.

The initial set of available libraries was acquired through a search in GitHub. Keywords and the type of the results are the two key limiting factors to guide the initial set of results. For the first limiting factor, the keywords "explainable AI" and "interpretability" were used. The second limiting factor was the type of results and this was set to "repository" which excluded all the results with these keywords in, e.g., the code itself or discussions, issues, commits, etc. Applying these limiting factors resulted in 57 results. To further narrow down the results, three rules were developed for the initial scan of the libraries as shown below:

1. Result has to be a repository of a Python library or a software package;
2. Result has to implement at least one xAI method;
3. Result has to be an overview repository (repository that provides an overview of xAI libaries).

Supplementary source code together with the overview of library versions and descriptions to recreate an exact development environment used for these experiments can be found on GitHub at the following URL: https://github.com/mathabaws/SOTA_xAI_Visual_analytics (accessed on 12 January 2022).

3. Results

3.1. Library Comparison on Glioma Subtype Classification

By using the processed data from the combined studies described in the materials section, we trained a model to classify cancer subtypes by distinguishing between the Oncotree codes GBM, AASTR, and ODG. These are the top three most frequent diffuse glioma subtypes samples.

In general, 1020 training instances were used for training, and 256 for testing. Testing data remained unbalanced representing a realistic scenario. Ten-fold cross-validation scored a mean accuracy of 0.87 with a standard deviation of 0.02. The results of the trained model are shown in Table 1.

Table 1. Predictive results using RF classifier.

	Random Forest Classifier			
Oncotree Code	**Precision**	**Recall**	**F1-Score**	**Support**
GBM	0.85	0.96	0.90	177
ODG	0.70	0.42	0.53	45
AASTR	0.90	0.79	0.84	34
macro avg	**0.82**	**0.73**	**0.76**	**256**

In the next subsections, the Python libraries suitable for xAI and VA selected for in-depth analysis are presented, including results from tests with the above described model.

3.2. Python Libraries for Explainability

Applying the method described in the previous section, 52 relevant repositories were identified. Moreover, several overview repositories in the topic of xAI have been identified. These overview repositories provided information on the libraries other than ones identified through initial scan and were further used for backward and forward search. Next, a process resembling abstract and conclusion scan was conducted to filter out the libraries not focused on xAI and/or VA. In other words, documentation from repositories and implementation of the libraries were scrutinized to identify their focus and scope. As a result, 48 libraries were selected as relevant. These libraries were analyzed, interpreted, and summarized in a concept-centric way [42]. Through an in-depth analysis, metadata was collected, and core libraries and frameworks were identified for further exploration. Figure 1 provides an overview of the process.

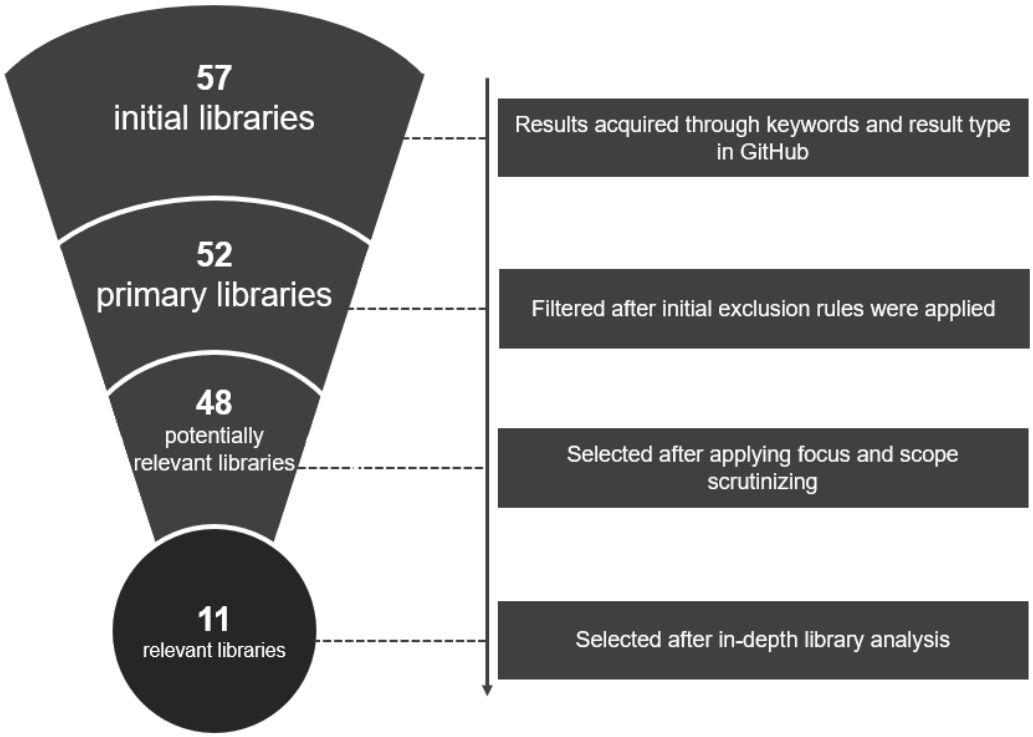

Figure 1. Overview of the review process.

As a first step, we drill down initial results described in the previous section to the most important libraries aiming for xAI using visualization tools. The complete comparison table can be found in Appendix A.1. We then defined structured rules that help us to identify relevant libraries, which will be further analyzed and experimented. Firstly, we select only those libraries that are implemented in Python and integrate visualization features to communicate xAI results. Furthermore, chosen libraries are able to explain classification models. Last but not least, these libraries are open source, provide good documentation, and support tabular data.

After filtering, we identified 11 relevant libraries. Selected libraries based on the aforementioned rules are listed in Table 2. We excluded 6 of the 11 identified libraries as missing criteria were revealed during the in-depth inspection. The remaining relevant libraries were grouped into three different groups: libraries aiming for global explainability in general, libraries focusing on local explanation, and, in particular, libraries which support Lime and SHAP approaches. In the first group, the following libraries are selected: ELI5 [43], Dalex [29], InterpretML [28], and SHAP [17]. In the second group, i.e., local explainability,

Lime and SHAP approaches are explored in more detail. Three different libraries focusing on Lime are analyzed: Lime [16], Dalex [29], and InterpretML. Finally, three different libraries focusing on SHAP approaches are analyzed in detail: InterpretML, Dalex, and SHAP. The selected libraries are analyzed and compared within the groups and the results are shown in the sections below. The complete overview table can be found on the GitHub repository (Appendix A.2). All experiments concerning the analyzed libraries in depth are conducted using a notebook with the following characteristics: Lenovo ThinkPad L470, Intel(R) Core(TM) 2.70GHz - 2.90GHz, 16 GB RAM, Windows 10.

Table 2. Summary containing library names and analyzed properties.

Library Name	Type of Explanation	Regression	Text	Images	Distributed	Licence
AI Explainability 360 (AIX360)	Local and Global	No	No	Yes	No	Apache 2.0
Alibi	Global explanation	Yes	No	No	No	Apache 2.0
Captum	Local and Global	Yes	Yes	Yes	Yes	BSD 3-Clause
Dalex	Local and Global	Yes	No	No	No	GPL v3.0
Eli5	Local and Global	Yes	Yes	Yes	No	MIT License
explainX	Local and Global	Yes	No	No	No	MIT License
LIME	Local and Global	No	Yes	Yes	-	BSD 2-Clause "Simplified" License
InterpretML	Local and Global	Yes	No	No	-	MIT License
SHAP	Local and Global	Yes	Yes	Yes	-	MIT License
TensorWatch	Local explanation	Yes	Yes	Yes	-	MIT License
tf-explain	Local explanation	Yes	Yes	Yes	-	MIT License

3.3. Global Explainability

Several libraries were identified with implementation of different feature importance methods. These are methods that rely on assigning a score to input features based on the predictive performance they add to the model. We are starting this overview with the focus on (1) methods for global explainability of the model and (2) methods that use visualization to communicate the explainability results. During the in-depth analysis, four libraries were identified to contain feature importance visualizations, namely ELI5, Dalex, InterpretML, and SHAP.

ELI5 focuses on feature selection with the implementation of permutation importance. It enables extraction and visualization of feature weights and their contribution from the model as a form of global explanations. Visualizations are based on the list view of the features and their weights in a tabular form. The gradient of green and red color indicates the positive or negative impact on the model decisions, and there are no interactive options. Figure 2 depicts feature importance visualization implemented in the ELI5 library. Furthermore, model inspection on the prediction level is supported, which uses similar visualization with weights adding up to either probability of a class in classification models or predicted value in case of regression models.

Dalex implements a method called variable importance which provides global explanations of a model based on Permutational Variable Importance [44]. Each variable is randomly shuffled in this method, and the model is inspected for its predictive performance. Intuitively, more important features impact the model performance more than the less important features. Finally, after 10 permutation rounds for each feature, visualization is created, showing the impact of each feature on the model. Such visualization provided by the Dalex library is depicted in Figure 3. Furthermore, the Dalex library provides a simple interactive overview during the mouse hovering over the visualization. This interactive window quantifies their influence on the model and provides additional information. The

Dalex library also provides the option to tune the hyperparameters, such as a number of permutation rounds and various thresholds, and enables grouping of the features.

Weight	Feature
0.1359 ± 0.2090	IDH1
0.1163 ± 0.1284	Overall Survival (Months)
0.1018 ± 0.0957	Diagnosis Age
0.0627 ± 0.1053	CIC
0.0586 ± 0.0460	Mutation Count
0.0456 ± 0.0549	TP53
0.0400 ± 0.0625	ATRX
0.0230 ± 0.0321	Overall Survival Status
0.0214 ± 0.0349	PTEN
0.0214 ± 0.0374	TERT
... *242 more* ...	

Figure 2. Visualization of feature weights and their impact to the model in the ELI5 library.

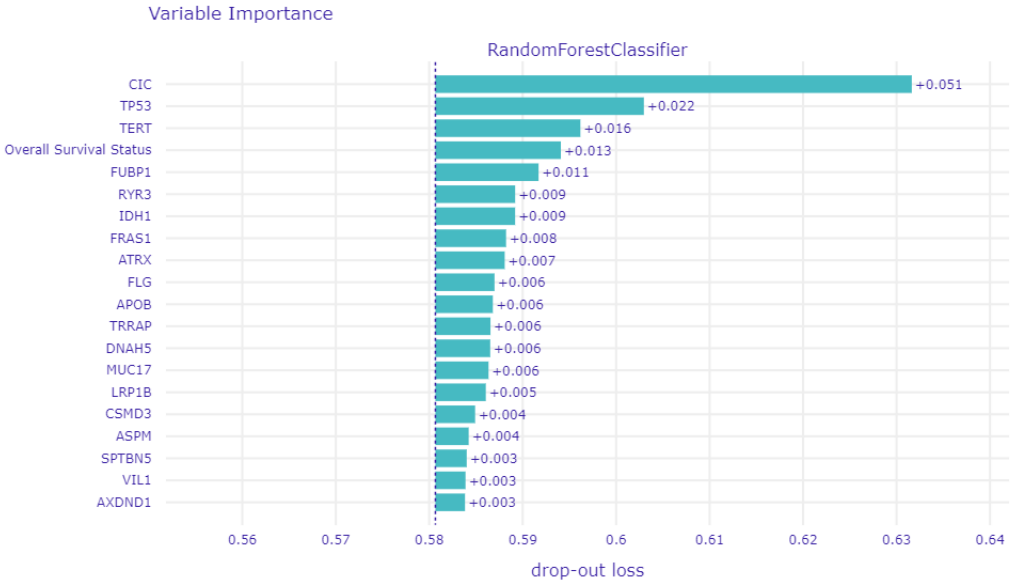

Figure 3. Visualization of permutational variable importance in the Dalex library.

The SHAP library provides the opportunity to analyze the model at the global level. This method helps to interpret the model by estimating feature importance altogether with feature effects on prediction with respect to raw data (as shown in Figure 4). The importance of features is shown along the x-axis, with important features listed at the top. For each feature, the contribution towards the specific classes is shown using the corresponding color, as shown in Figure 4a. Furthermore, SHAP provides the opportunity to conduct global interpretation for specific classes as shown in Figure 4b. In this case, the contribution of specific features is shown along x-Axis, where the contribution can be either positive (contributed toward prediction of this class) or negative. Each data point stacked vertically within this visualization represents the contribution for a specific instance. The color gradient encodes the raw values, blue representing the lowest and red the highest value.

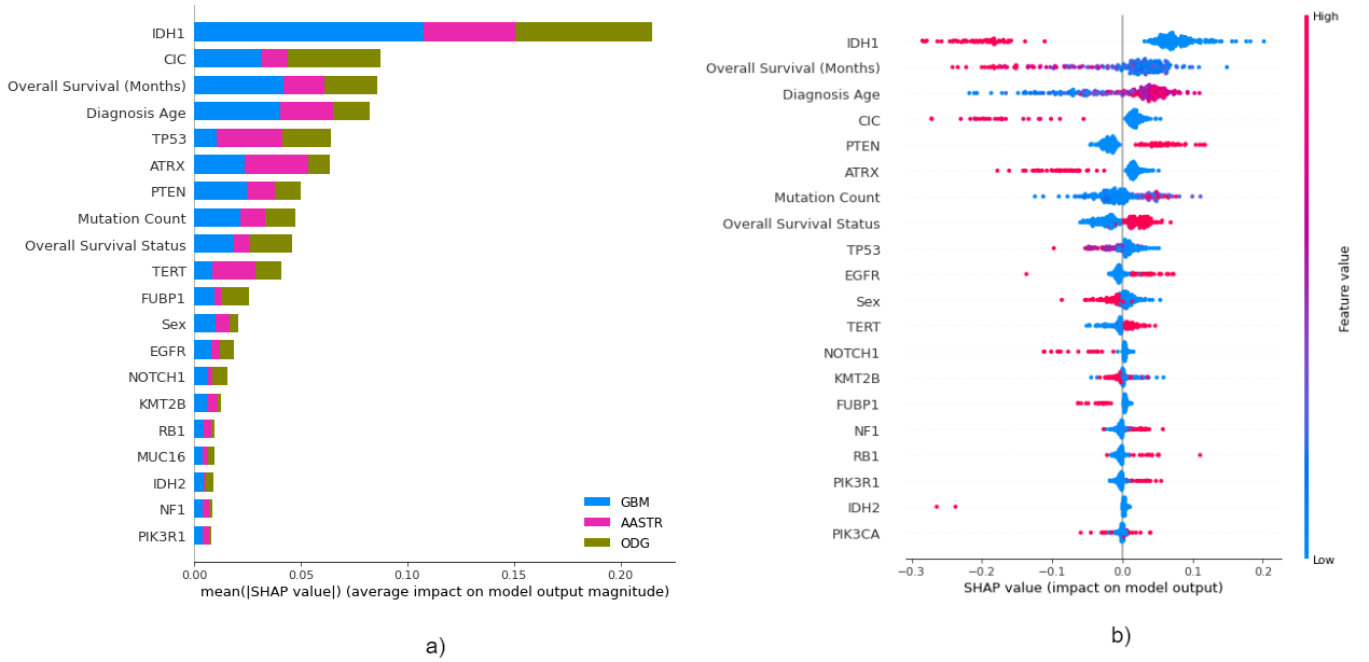

Figure 4. (**a**) Visualization of the impact of different variables in the global model performance in the SHAP library (**b**) Visualization (summary ploy) that combines feature importance with feature effects for a specific class (class "GBM" in this case).

As mentioned in Section 3.1, InterpretML is focused on navigation through different views and interactive application of different methods. One of the methods that is provided by the library is the overall importance. Overall importance presents the global feature importance of the model. InterpretML makes the distinction of algorithms that are applied in two different model types. These are glassbox models and black-box explainers. To be able to apply and extract global feature importance, a glassbox model needs to be trained. These models are structured for direct interpretability, contrary to the black-box models that provide approximations of explanations. This introduces additional overhead in utilizing InterpretML for model explainability, as an additional model had to be trained to extract important features of the model. An example of such feature importance visualization provided by InterpretML is depicted in Figure 5. Based on the popular visualization library Plotly [45], InterpretML allows simple interaction with the visualization (e.g., zoom-in, selection, export to image format, etc.).

Summarizing libraries for global explanation analysis, in terms of computational load, ELI5 provides the most lightweight solution for feature inspection. A simple and unified application programming interface enables a virtually instant overview of the features. On the contrary, all other remaining libraries require some degree of further processing to provide global explainability information. In the context of tabular data, the only supported visualization in ELI5 is a table overview with a gradient of green and red color encoding to indicate the importance of a feature in model predictions. The SHAP library provides more variety in terms of visualization with the implementation of bar chart and summary plot, which combines feature importance with feature effects. In regard to interactivity, visualizations provided by SHAP in the context of global importance are static and do not provide any further interactive features. Furthermore, in comparison to ELI5, SHAP requires an additional computational load that comes with the calculation of shap values. The Dalex library implements additional interactivity features in the model-level variable importance calculation. Visualization implemented in Dalex contains a list of features and their impact on predictions, with additional information provided upon the selection of a feature, which proved particularly useful when inspecting models with large numbers of features. However, this interactivity comes with additional computational load, which

was significant in comparison with other libraries. Calculation of the feature importance for the previously developed model took from 1.5 to 5 min, depending on the number of permutation rounds for each feature. Finally, InterpretML provided the most interactivity out of all previously described libraries. Invoking global explanation functions provided a menu system alongside visualizations to investigate feature importance and their interaction. Each visualization enabled extensive inspection through zoom, select, lasso, and export functionality. Despite this interactivity, limitations of InterpretML library are due to the requirement of using built-in GlassBox models such as ExplainableBoostingClassifier. Although showing comparable performance, this restriction to built-in models is quite significant. Furthermore, the additional computation overhead of training an additional model should not be overlooked. Overall, from the perspective of global explainability, all identified libraries provide useful insight into the model behavior, and each comes with its merits and limits from the perspective of visualization options, interactivity, and computational overhead.

Figure 5. Visualization of overview of feature importance provided by InterpretML.

3.4. Local Explainability

Models that produce accurate predictions and, at the same time, can explain such predictions are crucial. Researchers often generate global explanations, which try to explain predictions of black-box learning algorithms. However, such a global explanation cannot clarify the prediction of every single instance in the model. Local explainability focuses on gaining the user's trust for individual predictions and then trusting the model as a whole. Interpretation should make sense from the point of view of individual prediction. Globally important features may not be important locally and vice versa. In this case, the aim is to understand model decisions with respect to local context rather than the global behavior of the model.

There are several solutions mentioned in this paper and in this section; we will focus on the local explanations and two most relevant Python libraries, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) [16], identified by the selection rule mentioned in the previous section.

3.4.1. Local Explainability with SHAP

We identified three different libraries that fit to the selection rule of the most relevant libraries which are implementations of the SHAP approach: InterpretML [28], Dalex [29], and SHAP [17]. Consequently, we compared and analyzed these libraries showing the state-of-the-art in the topic of SHAP values aiming for the interpretation of black-box models.

Dalex (shown in Figure 6) offers basic interaction such as hovering over the visualization. This provides an opportunity to navigate through the results easily. Moreover, it provides the option to download the chart directly from generated visualization.

SHAP offers various visualization such as waterfall graphs for global analysis and force plots for local analysis. We specifically compared local interpretation based on the force plot shown in Figures 7 and 8. SHAP provides many alternatives to interpret black-box behaviors, such as the force plot of a single prediction shown in Figure 7, which is a static visualization. Additionally, in Figure 8 a grouped analysis of all predicted instances is shown, where the single instances are stacked over the x-axis. This interactive visualization provides the opportunity to select a method (e.g., ordered by similarity) to order the instances over the x-axis group the results using the drop-down menu on the top of the chart over the x-axis. Moreover, on the y-axis, the drop-down menu offers the option to select the feature which the user wants to analyze. Moreover, hovering over the chart highlights different details, thus increasing the level of information provided from this approach.

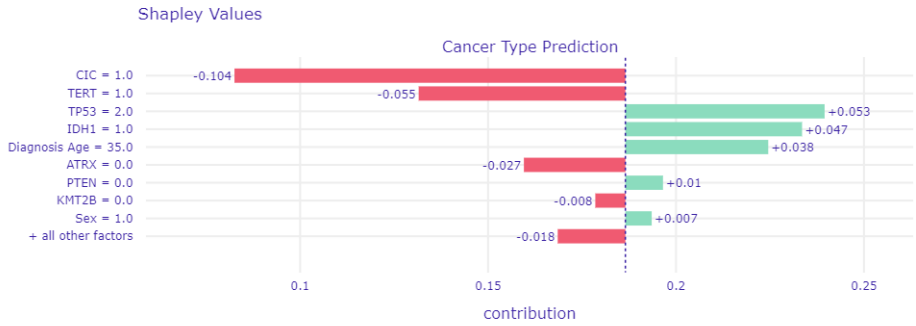

Figure 6. Visual explanation of black-box prediction results using the Dalex library. In this case, an exemplary local view of class GBM is detailed.

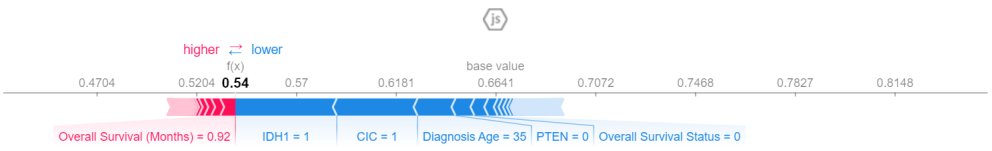

Figure 7. Local visual explanation of black-box prediction results (exemplary instance of class GBM) using the SHAP library.

In contrast, InterpretML provides an opportunity to navigate through different instances using a drop-down menu, presented in Figure 9. The estimated SHAP results for the specific instance are shown automatically by selecting a particular instance. This provides an opportunity to navigate through different instances, having a better overview of the results and the possibility to compare the output of different instances faster. In particular, information such as the predicted class, actual class, and residual error for each instance is shown in the drop-down menu, as well as in the main window. This provides an opportunity to compare similar instances based on predicted class, actual class, or the residual error, thus showing an opportunity to understand a model's class prediction more comprehensively. Moreover, interactions such as zoom in, zoom out, pan, select, and download are supported. However, InterpretML supports only KernelSHAP methods.

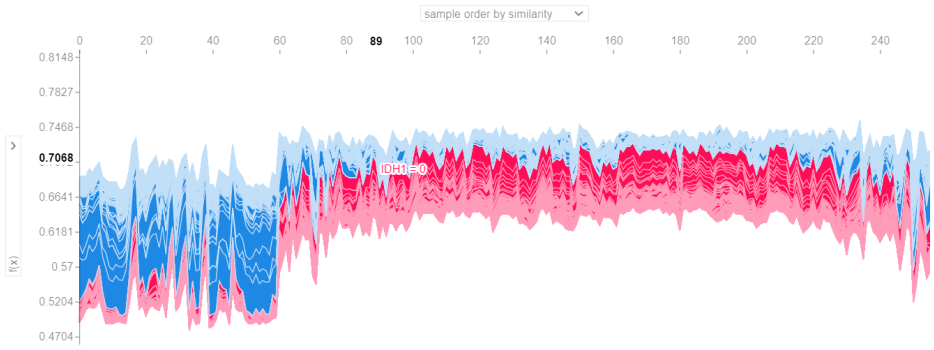

Figure 8. Grouped based analysis of instance prediction. Local visual explanation of black-box prediction results using the SHAP library. In the x-axis, the local explanation results of every instance are stacked. The y-axis shows the contribution to prediction and the option to select the feature that will be explored for every instance in terms of SHAP contribution.

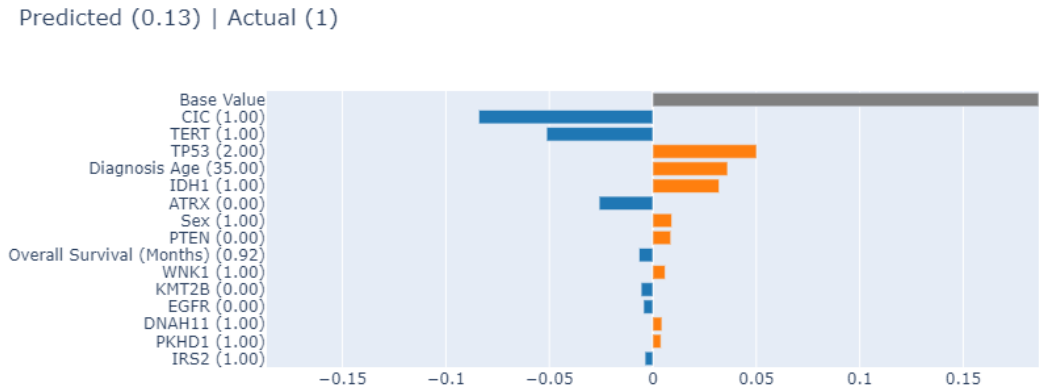

Figure 9. Black-box model interpretation using InterpretML. In this case, an exemplary local view of class GBM is explored in detail.

Although InterpretML provides multiple interaction possibilities to explore the black-box model, it still presents the highest computation overload. InterpretML takes approximately 95.4 s modeling time per instance and 0.73 s visualization time per instance. SHAP requires 21.2 s modeling time and approximately 0.15 s visualization time for single instances charts and 1.03 s for grouped instances plots. Finally, Dalex needs fewer computation resources with around 0.143 s modeling time and 1 m and 49 s visualization time.

3.4.2. Local Explainability with LIME

LIME (Local Interpretable Model-Agnostic Explanations) is a popular technique that tries to explain the predictions of any classifier by learning an interpretable model locally around the prediction. The key idea behind LIME is that it is easier to approximate a black-box model by a simple model locally. The Lime library can explain any black-box classifier with two or more classes. The visualization output of the LIME library is a list of explanations, reflecting the contribution of each feature to the instance prediction (Figure 10a). Visualization provides local explainability and helps to investigate which feature changes will have the most impact on the instance prediction.

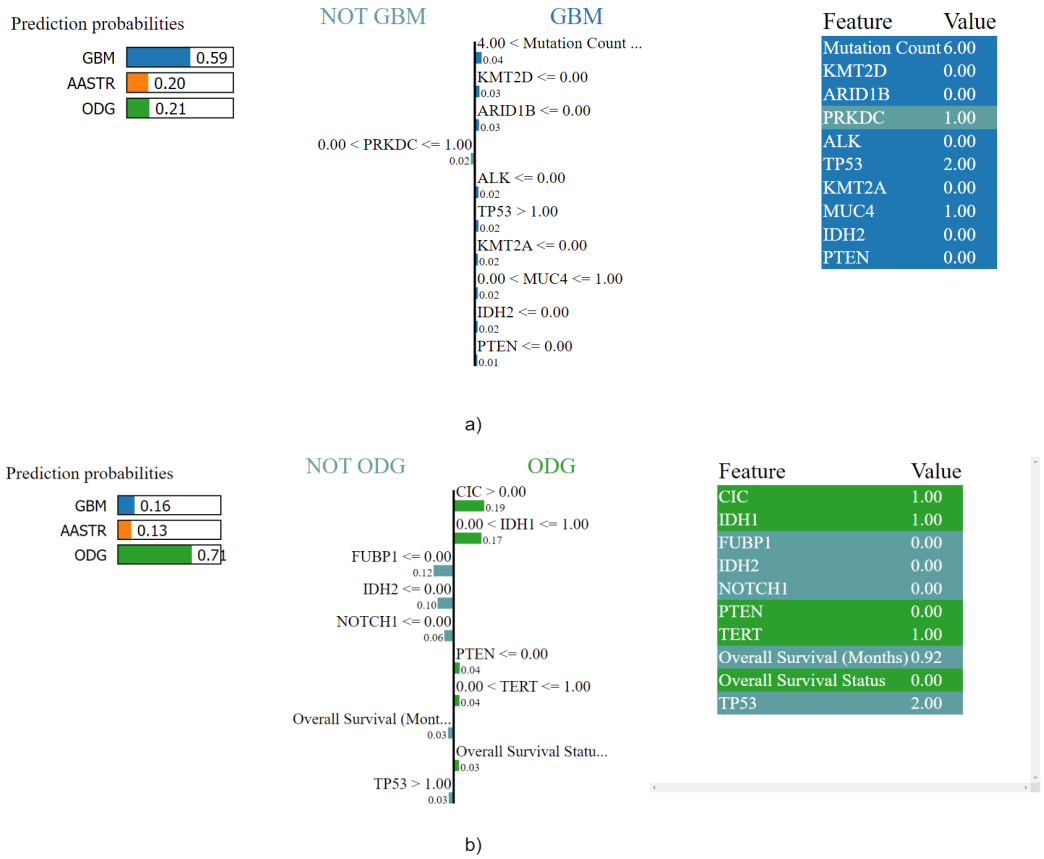

Figure 10. Local visual explanation of black-box prediction results using LIME library—exemplary instance prediction of class (**a**) GBM and (**b**) ODG.

Figure 10a,b show instance explanations of LIME. These figures provide explanations for an instance prediction on the class of GBM or ODG, respectively. There are three parts of LIME visualization: a class description with an accurate prediction for each class, a plot showing the impact of features, and a table with actual values in the instance. The left-most section displays prediction probabilities. For the multi-class classification task, we have three colors, blue (GBM), orange (AASTR), and green (ODG). The middle section returns the most important features. The impact of features helps the user to understand which features values are supporting class prediction positively (right side) and which features values are not supporting prediction (left side). If we take Figure 10a as an example, features are represented in two colors: blue and light sea-green. The blue bars indicate supporting (positive) scores towards an instance being predicted as GBM, while the light sea-green bar indicates contradicting (negative) scores towards its prediction. Float point numbers on the horizontal bars represent the relative importance of these features. We can see in Figure 10a that the highest positive influence have genes CIC, BCL6, PKD1L1, and ATRX.

Similar to the SHAP approach, besides LIME, InterpretML and Dalex are the most relevant libraries that implement the LIME approach, based on our selection rule. The libraries Dalex and InterpretML were already mentioned and explained in previous sections. The resulting plot for Dalex is shown in Figure 11. The Figure shows an explanation for instance predicted as class GBM. The length of the bar indicates the magnitude, while the color indicates the sign (red for negative, green for positive) of the estimated coefficient. In the previous examples, Dalex offered basic interaction such as hovering over the visualization, as well as the ability to navigate through the results easily. Unfortunately, the resulting plots for the LIME method do not provide any of these features.

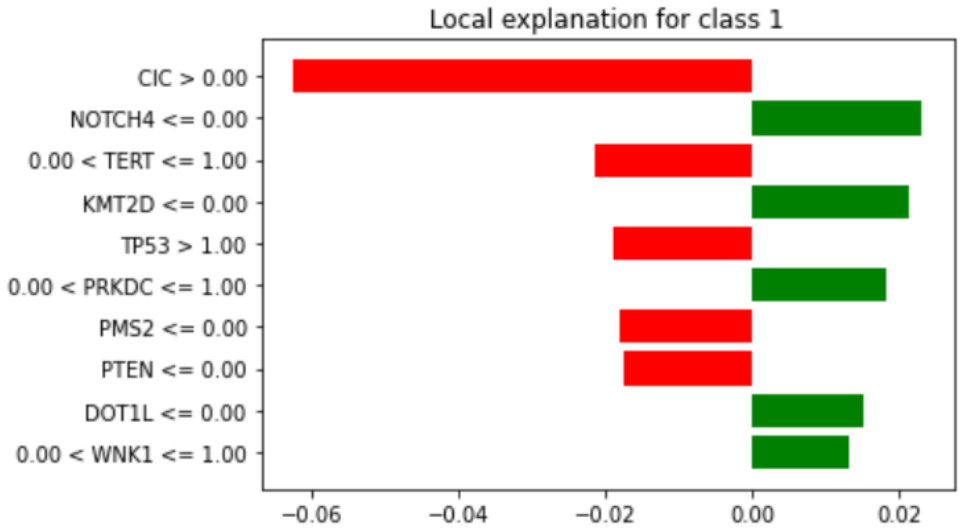

Figure 11. Visual explanation of black-box prediction results using the Dalex library (class GBM)—LIME approach.

InterpretML using the LIME approach is shown in Figure 12. As in previous examples (see Figure 9), InterpretML provides an opportunity to navigate through different instances using a drop-down menu. By selecting a specific instance, we can navigate through different instances having a better overview of the results.

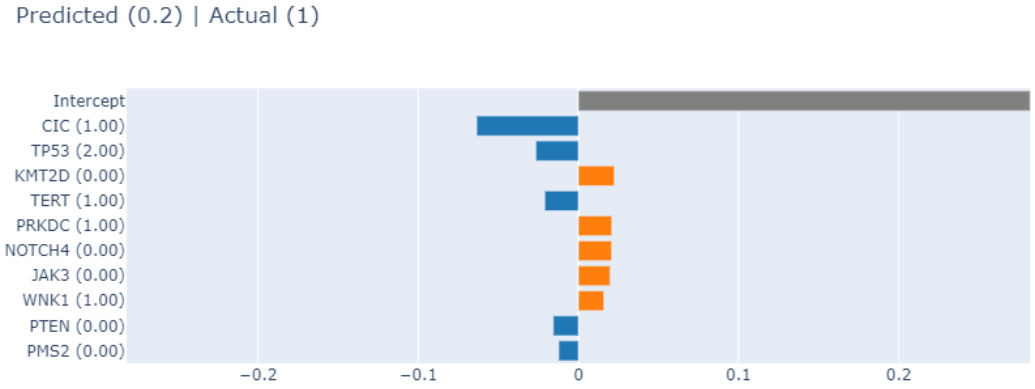

Figure 12. Visual interpretability of black-box model using the InterpretML library with the LIME approach. In this case, the local view of class GBM is explored in detail. Colors are encoded as follows: blue = negative contribution, orange = positive contribution, and gray = intercept.

Regarding computation time, as can be seen in Table 3, InterpretML presents the highest computation overload. InterpretML takes approximately 7.28 s modeling time per instance and 0.72 s visualization time. LIME requires 3.63 s modeling time and 0.4 s visualization time. Finally, Dalex needs little bit more computation resources, with around 3.97 s modeling time and 0.78 s visualization time.

Table 3. Library comparison with respect to global and local explainability.

Library	Computation Overload—Modeling	Computation Overload—Visualization	Interactivity
Global Explainability			
ELI5	-	0.19 s	not interactive (5)
Dalex	1 m 20.07 s	0.33 s	slightly interactive (4)
SHAP	13.21 s	0.32 s	not interactive (5)
InterpretML	7.91 s	9.37 s	very interactive (1)
Local Explainability—SHAP			
SHAP	21.2 s	0.15 s	very interactive (1)
InterpretML	95.4 s	0.73 s	very interactive (1)
Dalex	0.143 s	1 m and 49 s	interactive (3)
Local Explainability—LIME			
Lime	3.63 s	0.4 s	not interactive (5)
InterpretML	7.28 s	0.72 s	very interactive (1)
Dalex	3.97 s	0.78 s	not interactive (5)

3.5. Biomedical Implication of Features

The evaluation of features affecting the classification between the diffuse glioma (DIFG) of Glioblastoma multiforme (GBM), Anaplastic Astrocytoma (AASTR), and Oligodendroglioma (ODG) highlights various mutated genes and clinical variables depending on the underlying xAI method. Diagnosis age and survival are among the most important predictors all of the methods, followed by varying gene mutations. Capicua (CIC) depicts an important feature in all approaches and is the most valuable gene feature in Dalex, second in SHAP and InterpretML, and fourth in ELI5. Mutated IDH1 is among the top features and, from a clinical point of view, commonly used for survival prognosis in patients suffering from glioma [34]. Further important variables highlighted by the different xAI methods in different order also include other biomarkers used for clinical classification of glioma, such as ATRX, TP53, TERT, PTEN, or EGFR. Local explanations show a partly different picture and detail decisions of the algorithms on local examples. We present Figures on local instances on the class of GBM (Figures 7, 9–11) and (b) ODG (Figure 10). Variables changed place in the hierarchy of importance, while there is additional information on a particular variable's prediction impact shown as negative or positive factor towards the particular class of the local view.

3.6. Overview of xAI Approaches

The comparison overview and ranking is shown in Table 3. As a result, the table shows the overview concerning the global and local explainability comparison results of SHAP and LIME.

In the context of global explainability, similar criteria can be used for the selection of libraries, i.e., computational overhead, implemented visualizations, and interactivity. From the perspective of the computational overhead, ELI5 provides the most lightweight solution both in terms of computational overhead and implemented visualizations and interactivity. The simple interface provides a good basis for a quick inspection of the existing model and overall model debugging. Feature importance alongside other implemented functionality (e.g., feature selection) of ELI5 can be convenient during the model development process. Increased interactivity and visualization options come with the additional computational overhead in SHAP, Dalex, and InterpretML libraries. From the perspective of interactivity in global explainability, InterpretML provides the most interactive solution. The addition of menu components to select different model components makes it easy to switch between

analysis perspectives and extensive visualization features (zoom, lasso, select, and others). This provides excellent analytical insights. However, these functionalities come with limitations in terms of the limited scope of built-in Glassbox models that can be used and additional computation overhead caused by model retraining. In terms of visualization, SHAP and Dalex are in between ELI5 and InterpretML. Compared to ELI5, Dalex requires more computational overhead but provides additional interactivity and visualizations. On the other hand, SHAP requires even more computational overhead but provides excellent visualization options that enable a complex analysis of the interplay between feature importance and feature effect. From the perspective of the stage of the development of the predictive model, ELI5 and Dalex seem to be focused on the model analysis, while SHAP and InterpretML put focus on the underlying data and how this data impacts the model decisions.

Regarding local explainability using the SHAP approach, we identified different outcomes. In general, to explain a black-box in the big data context, it is important to find the trade-off between computation resources and explainable results. In the context of local explainability, SHAP outperformed other libraries in terms of computational resources and providing an interactive way to explore the different model predictions. In terms of interactivity, both SHAP and InterpretML outperform Dalex and provide many options to analyze explainable results of multiple instances interactively. However, if the goal is to find a trade-off between computational overhead and interactivity, then Dalex seems as the optimal solution in this context. Finally, if the focus is on exploring the features, the SHAP force plot grouping methods provide many advantages. However, InterpretML offers the option to compare different instances in terms of feature contribution, predicted class, actual class, and residual error. This provides a huge advantage over other methods for analyzing the behavior of block box models in terms of predicted/actual class. Compared to SHAP, LIME has advantages in terms of speed as it builds the model around individual predictions. In the case of large datasets, using SHAP might not be feasible due to the large computational overhead caused by the calculation of all global permutations. Despite the performance overhead, SHAP provides a unified solution, which, once computed, offers more refined explainability and analytical experience.

LIME provides an intuitive instance explanation. The LIME library builds the model around individual predictions (neighborhood), thus it does not take additional time to compute the model for all instances. On the other hand, the resulting plots do not provide any interactivity. Using Dalex for the LIME approach does not offer any interaction as for the other libraries. InterpretML is the only library providing interactivity while using the LIME approach. In comparison with the LIME plot, InterpretML's resulting plot does not offer an extensive summary of features.

The main advantage of SHAP for local explanation is that it is the only xAI method based on solid theory (Shapely value) [46]. Moreover, SHAP guarantees that the prediction is fairly distributed among all feature values. On the other hand, LIME for local explanation is faster than SHAP concerning computation time. In particular, if the aim is to analyze huge data sets, then LIME will provide a suitable alternative to the time-consuming computation of Shapely values. The SHAP approach considers this challenge by using approximation and optimization; however, not all model types are supported yet. In particular, LIME supports tabular data, text, and images. In other xAI methods, it is rare that all these types of data are supported.

4. Discussion

The output of any ML model should be comparable and interpretable. This is of particular interest to researchers in the medical domain as for cancer, where model performance may be compared with the one of clinicians [47]. Some experts from the medical domain argue that transparency for black boxes is not of primary interest to AI applications in their domain, as doctors make diagnoses based on their experience, and complete information on the causality of medical issues are rare [48,49]. However, xAI methods can help to

gain new insights and forward biomedical knowledge to better understand interrelated characteristics and signaling components in pathologies.

As a modeling approach, classifying glioma sub-types is exemplified: As the chosen dataset combining data from different brain tumor studies comprises sample data primarily from the glioma subtypes GBM, AASTR, and ODG, these three disease types were chosen to be classified to apply VA methods for interpreting global as well as local feature importance. The dataset provides Oncotreecode as identifier. GBM, AASTR, and ODG are all DIFG subtypes. Even combining data from six different studies resulted in a lack of samples for specific subtypes, therefore only the top three were chosen. Open data resources are still set to develop further and to be extended [50]. The chosen dataset is unbalanced and fits this use case insofar as it represents an often-found challenge in molecular sciences. This study aims to describe xAI tools rather than to provide a highly performing classifier solution; still, classifying glioma subtypes is a challenging task, which makes it an ideal example for comparing VA features in xAI. Cross-validation of xAI is not applicable to date, as a matter of ongoing research.

From a biomedical point of view, many of the important variables highlighted by the various xAI methods are already known to be involved in cancer signaling and represent common biomarkers in glioma. Generally, such insights into the model can be used for validation. The transcriptional repressor CIC is part of the tyrosine kinase signaling pathway which is known to be involved in tumorigenesis, especially in GBM [51]. Other gene features impacting the classification include mutated IDH1, ATRX, TP53, PTEN, TERT, NF1, and EGFR, all of which are known to be involved in DIFG [22,52]. Among important variables are also the mucin protein family (MUC16 and MUC17) which are involved in epithelial barrier formation and potential biomarkers for favorable prognosis in DIFG, or lysine methyl transferase (KMT2B) also shown to be a player in gliomagenesis [53,54]. One example given, the type I transmembrane protein Notch 1 receptor (Notch1), is involved in the NF-κB signaling pathway effecting cancer development and progression, especially in GBM [55]. Notch 1 is listed in the global top 20 variables listed by SHAP, but not by Dalex. Still, in SHAP it distinguishes primarily between ODG and GBM. Some gene mutations are not primarily common for one class of sub disease, but can increase or mitigate cancer malignancy as given by the example of IDH1. Mutated IDH1 will lead to a favorable outcome, but a complete genetic profile could tell more of cases not concordant with standard prognoses [56]. In the case of local explanations as given in Figure 10b, IDH1 is selected in favor of the ODG class. Local explanations can thereby support further insight on individual cases instead of presenting the big picture of global classes.

The local explanation in Figure 12 shows that the low mutation count has been used to select for the class of GBM for this instance. A high mutational burden is indicative for an unfavorable prognosis as given by GBM, which would contradict the observation in this local view. This could be seen as a limitation of model accuracy or be used for future investigations on individual cases and underlying experimental constraints. In Figure 10, we can see another local explanation for GBM classification which is supported by low numbers of mutation count. This could be due to the fact that a high number of samples originate from GBM biopsies, so that samples with low mutation count can also be frequently found. This unbalanced data source can be seen as a certain limitation to the represented model; however, combining local explanations in Figure 10 with global explanations in Figure 4, we can see that even if the mutation count is among the top rated features, there are also other important features that should be taken into account for further analysis. Diagnosis age and overall survival are preferably incorporated by the different algorithms on a global basis. Further local instances by InterpretML and Dalex are presented in Figures 9 and 11. For example, gene mutations With-No-Lysine Kinase 1 (WNK1) are ranked among the top important features, highlighting a possible role of WNK1 in glioma, which has yet to be shown for WNK3 [57]. One local instance presented by Lime in Figure 10a ranks AT-Rich Interaction Domain 1B (ARID1B), shown as putative driver gene in glioma [58], among the most important variables for classifying

GBM. The feature is followed by others such as Protein Kinase DNA-Activated Catalytic Subunit (PRKDC), a component of the autophagy-regulating signaling cascades to be alterated also in glioma [59], and the Anaplastic Lymphoma Receptor Tyrosine Kinase (ALK), whose variation has been implicated with pediatric glioma [60]. Another local instance by InterpretML, shown in Figure 9 includes Polycystic Kidney And Hepatic Disease 1 Protein (PKHD1), shown as variant in GBM [61], in the top feature list, followed by Insulin Receptor Substrate 2 (IRS2) [62] and Dynein Axonemal Heavy Chain 11 (DNAH11), which has been recently linked to immune cell infiltration in glioma [63].

Applying xAI methods further facilitates the refinement process of the model's underlying data and thereby helps to understand and enhance a model. By studying the results of local explainability methods, we found an error in the algorithm for computing the different gene's mutations. The value "NA" had been counted as 1 rather than 0, due to the fact that different gene mutations from the processed data are handled as strings, separated by empty spaces. After evaluating and comparing the results, we corrected the model and revisited the comparison, leading to better results, both in reproducibility of already known markers and better quality, as well as model performance.

The comparison of xAI libraries can be used for gaining biomedical insights, but also to detail advantages and challenges using these tools appropriate for certain application scenarios. Figures 8 and 11 show two diverging examples in VA feature range such as interactivity or details on demand regarding xAI quality and quantity. After all, which library and approach to choose depends on the use case, such as finding novel biomarkers in analyzing classification feature importance or investigating survival prediction. Therefore, we compared libraries regarding their global xAI features separately from those with local ones. By making use of the detailed descriptions above, we try to support the decision-making process of choosing a suitable library. F.i. ELI5 is optimal regarding computational load, while InterpretML offers most interactivity at the expense of computation time.

5. Conclusions

We present a comparison of the ease of use of current xAI libraries and exemplify how to support understanding of a black-box model's results in glioma classification to find novel biomarkers. Thereby, we describe possibilities how to integrate VA features for xAI. We only scratch the surface when it comes to going beyond xAI. The process of understanding can be supported by interactivity and other features to assess the quality of explanations [64]. Future work may also include taking the type of mutation into account by incorporating various types of mutations as different features—for now, the model differentiates between wild-type/mutated and number of mutation if there is more than one mutation for the same gene. Additionally, data could be integrated from miscellaneous sources and cover further subclasses or clinical features, while adding use cases of survival prediction or clustering approaches for signaling insights. Performance experiments for further information on requirements and recommendations could be also part of future work. Finally, we believe that the presented approach, using open data, providing open source implementation, and focusing on ease of use, as well as showcasing the application of xAI to real scientific problems, can contribute to the research fields of cancer science and beyond.

Author Contributions: Conceptualization, C.J.-Q. and F.J.; methodology, all authors; software, M.V., N.J. and M.G.; validation, F.J.; formal analysis, C.J.-Q.; investigation, all authors; resources, all authors; data curation, C.J.-Q., F.J. and M.V.; writing—original draft preparation, all authors; writing—review and editing, F.J., C.J.-Q., A.H., S.T. and M.G.; visualization, M.V., N.J. and M.G.; and supervision, C.J.-Q. and F.J.; All authors have read and agreed to the published version of the manuscript.

Funding: Parts of this work have been funded by the Austrian Science Fund (FWF), Project: P-32554 "A reference model of explainable Artificial Intelligence for the Medical Domain". Additionally, two authors have been partially supported by the FFG, Contract No. 854184: "Pro²Future is funded within the Austrian COMET Program Competence Centers for Excellent Technologies under the auspices of the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and

Technology, the Austrian Federal Ministry for Digital and Economic Affairs and of the Provinces of Upper Austria and Styria. COMET is managed by the Austrian Research Promotion Agency FFG".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Preprocessed data and implementations such as notebooks can be found on https://github.com/mathabaws/SOTA_xAI_Visual_analytics/ (accessed on 12 January 2022).

Acknowledgments: We thank the cBioPortal maintainers and collaborators for providing data on cancer and all the other data providers to make open science possible. We dedicate our work in memoriam to our family members and friends we have lost.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AASTR	Anaplastic Astrocytoma
DIFG	Diffuse Glioma
CIC	Capicua gene
GBM	Glioblastoma multiforme
LIME	Local Interpretable Model-Agnostic Explanations
ODG	Oligodendroglioma
SHAP	SHapley Additive exPlanations
VA	Visual Analytics
xAI	explainable Artificial Intelligence

Appendix A

Appendix A.1. Complete Table of All Identified xAI Libraries

The full table listing all search results and filter criteria for comparing explainable libraries can be found via https://github.com/mathabaws/SOTA_xAI_Visual_analytics/tree/main/data (accessed on 12 January 2022).

Appendix A.2. Implementation Details

The repository containing code and experiments can be found via https://github.com/mathabaws/SOTA_xAI_Visual_analytics (accessed on 12 January 2022).

References

1. Bhardwaj, R.; Nambiar, A.R.; Dutta, D. A study of machine learning in healthcare. In Proceedings of the 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, Italy, 4–8 July 2017; Volume 2, pp. 236–241.
2. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]
3. Galon, J.; Pagès, F.; Marincola, F.M.; Angell, H.K.; Thurin, M.; Lugli, A.; Zlobec, I.; Berger, A.; Bifulco, C.; Botti, G.; et al. Cancer classification using the Immunoscore: A worldwide task force. *J. Transl. Med.* **2012**, *10*, 1–10. [CrossRef] [PubMed]
4. Murtaza, G.; Shuib, L.; Abdul Wahab, A.W.; Mujtaba, G.; Nweke, H.F.; Al-garadi, M.A.; Zulfiqar, F.; Raza, G.; Azmi, N.A. Deep learning-based breast cancer classification through medical imaging modalities: State of the art and research challenges. *Artif. Intell. Rev.* **2020**, *53*, 1655–1720. [CrossRef]
5. Carrio, A.; Sampedro, C.; Rodriguez-Ramos, A.; Campoy, P. A review of deep learning methods and applications for unmanned aerial vehicles. *J. Sens.* **2017**, *2017*, 3296874. [CrossRef]
6. Razzak, M.I.; Naz, S.; Zaib, A. Deep learning for medical image processing: Overview, challenges and the future. In *Classification in BioApps*; Springer: Cham, Switzerland, 2018; pp. 323–350.
7. Vukovi´c, M.; Thalmann, S. Causal Discovery in Manufacturing: A Structured Literature Review. *J. Manuf. Mater. Process* **2022**, *6*, 10. [CrossRef]
8. Gashi, M.; Ofner, P.; Ennsbrunner, H.; Thalmann, S. Dealing with missing usage data in defect prediction: A case study of a welding supplier. *Comput. Ind.* **2021**, *132*, 103505. [CrossRef]

9. Holzinger, A.; Kieseberg, P.; Weippl, E.; Tjoa, A.M. Current Advances, Trends and Challenges of Machine Learning and Knowledge Extraction: From Machine Learning to Explainable AI. In *Machine Learning and Knowledge Extraction*; Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E., Eds.; Springer International Publishing: Cham, Siwitzerland, 2018; pp. 1–8.
10. Holzinger, A.; Goebel, R.; Mengel, M.; Müller, H. *Artificial Intelligence and Machine Learning for Digital Pathology: State-of-the-Art and Future Challenges*; Springer Nature: Cham, Switzerland, 2020; Volume 12090.
11. Adadi, A.; Berrada, M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [CrossRef]
12. Castelvecchi, D. Can we open the black box of AI? *Nat. News* **2016**, *538*, 20. [CrossRef]
13. Samek, W.; Montavon, G.; Vedaldi, A.; Hansen, L.K.; Müller, K.R. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Springer Nature: Cham, Switzerland, 2019; Volume 11700.
14. Königstorfer, F.; Thalmann, S. Software documentation is not enough! Requirements for the documentation of AI. *Digit. Policy Regul. Gov.* **2021**, *23*, 475–488. [CrossRef]
15. Polzer, A.; Fleiß, J.; Ebner, T.; Kainz, P.; Koeth, C.; Thalmann, S. Validation of AI-based Information Systems for Sensitive Use Cases: Using an XAI Approach in Pharmaceutical Engineering. In Proceedings of the 55th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2022.
16. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
17. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 4765–4774.
18. Holzinger, A.; Biemann, C.; Pattichis, C.S.; Kell, D.B. What do we need to build explainable AI systems for the medical domain? *arXiv* **2017**, arXiv:1712.09923.
19. Katuwal, G.J.; Chen, R. Machine learning model interpretability for precision medicine. *arXiv* **2016**, arXiv:1610.09045.
20. Jiarpakdee, J.; Tantithamthavorn, C.; Dam, H.K.; Grundy, J. An empirical study of model-agnostic techniques for defect prediction models. *IEEE Trans. Softw. Eng.* **2020**, *48*, 166–185. [CrossRef]
21. Tan, S.; Caruana, R.; Hooker, G.; Lou, Y. Detecting bias in black-box models using transparent model distillation. *arXiv* **2017**, arXiv:1710.06169.
22. Jean-Quartier, C.; Jeanquartier, F.; Ridvan, A.; Kargl, M.; Mirza, T.; Stangl, T.; Markaĉ, R.; Jurada, M.; Holzinger, A. Mutation-based clustering and classification analysis reveals distinctive age groups and age-related biomarkers for glioma. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 1–14. [CrossRef] [PubMed]
23. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 2048–2057.
24. Keim, D.A.; Mansmann, F.; Stoffel, A.; Ziegler, H. Visual analytics. In *Encyclopedia of Database Systems*; Springer: Berlin/Heidelberg, Germany, 2009.
25. Samek, W.; Wiegand, T.; Müller, K.R. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv* **2017**, arXiv:1708.08296.
26. Gashi, M.; Mutlu, B.; Suschnigg, J.; Ofner, P.; Pichler, S.; Schreck, T. Interactive Visual Exploration of defect prediction in industrial setting through explainable models based on SHAP values. In Proceedings of the IEEE InfoVIS 2020 , Virtuell, MZ, USA, 25–30 October 2020.
27. Spinner, T.; Schlegel, U.; Schäfer, H.; El-Assady, M. explAIner: A visual analytics framework for interactive and explainable machine learning. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 1064–1074. [CrossRef] [PubMed]
28. Nori, H.; Jenkins, S.; Koch, P.; Caruana, R. InterpretML: A Unified Framework for Machine Learning Interpretability. *arXiv* **2019**, arXiv:1909.09223.
29. Baniecki, H.; Kretowicz, W.; Piatyszek, P.; Wisniewski, J.; Biecek, P. Dalex: Responsible Machine Learning with Interactive Explainability and Fairness in Python. *arXiv* **2020**, arXiv:2012.14406.
30. Li, X.H.; Cao, C.C.; Shi, Y.; Bai, W.; Gao, H.; Qiu, L.; Wang, C.; Gao, Y.; Zhang, S.; Xue, X.; et al. A survey of data-driven and knowledge-aware explainable AI. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 29–49. [CrossRef]
31. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A review of machine learning interpretability methods. *Entropy* **2021**, *23*, 18. [CrossRef] [PubMed]
32. Vilone, G.; Longo, L. Explainable Artificial Intelligence: A Systematic Review. *arXiv* **2020**, arXiv:2006.00093.
33. Masui, K.; Mischel, P.S.; Reifenberger, G. Molecular classification of gliomas. *Handb. Clin. Neurol.* **2016**, *134*, 97–120. [PubMed]
34. Louis, D.N.; Perry, A.; Wesseling, P.; Brat, D.J.; Cree, I.A.; Figarella-Branger, D.; Hawkins, C.; Ng, H.; Pfister, S.M.; Reifenberger, G.; et al. The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro-Oncology* **2021**, *23*, 1231–1251. [CrossRef] [PubMed]
35. Kundra, R.; Zhang, H.; Sheridan, R.; Sirintrapun, S.J.; Wang, A.; Ochoa, A.; Wilson, M.; Gross, B.; Sun, Y.; Madupuri, R.; et al. OncoTree: A cancer classification system for precision oncology. *JCO Clin. Cancer Inform.* **2021**, *5*, 221–230. [CrossRef]

36. Komori, T. Grading of adult diffuse gliomas according to the 2021 WHO Classification of Tumors of the Central Nervous System. *Lab. Investig.* **2021**, *67*, 1–8. [CrossRef]
37. Zacher, A.; Kaulich, K.; Stepanow, S.; Wolter, M.; Köhrer, K.; Felsberg, J.; Malzkorn, B.; Reifenberger, G. Molecular diagnostics of gliomas using next generation sequencing of a glioma-tailored gene panel. *Brain Pathol.* **2017**, *27*, 146–159. [CrossRef]
38. Van Lent, M.; Fisher, W.; Mancuso, M. *An Explainable Artificial Intelligence System for Small-Unit Tactical Behavior*; AAAI Press: Palo Alto, CA, USA, 1994; pp. 900–907.
39. Shin, D.; Park, Y.J. Role of fairness, accountability, and transparency in algorithmic affordance. *Comput. Hum. Behav.* **2019**, *98*, 277–284. [CrossRef]
40. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401–404. [CrossRef]
41. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **2013**, *6*, pl1. [CrossRef]
42. Webster, J.; Watson, R.T. Analyzing the past to prepare for the future: Writing a literature review. *MIS Q.* **2002**, *26*, xiii–xxiii.
43. ELI5's Documentation. Available online: https://eli5.readthedocs.io/en/latest/overview.html (accessed on 12 January 2022).
44. Fisher, A.; Rudin, C.; Dominici, F. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 1–81.
45. Databricks. *Collaborative Data Science*; Databricks: San Francisco, CA, USA, 2015.
46. Shapley, L.S.; Kuhn, H.; Tucker, A. Contributions to the Theory of Games. *Ann. Math. Stud.* **1953**, *28*, 307–317.
47. Kleppe, A.; Skrede, O.J.; De Raedt, S.; Liestøl, K.; Kerr, D.J.; Danielsen, H.E. Designing deep learning studies in cancer diagnostics. *Nat. Rev. Cancer* **2021**, *21*, 199–211. [CrossRef]
48. McCoy, L.G.; Brenna, C.T.; Chen, S.S.; Vold, K.; Das, S. Believing in black boxes: Machine learning for healthcare does not need explainability to be evidence-based. *J. Clin. Epidemiol.* **2021**. [CrossRef]
49. Wang, F.; Kaushal, R.; Khullar, D. Should health care demand interpretable artificial intelligence or accept "black box" medicine? *Lab Investig.* **2020**. [CrossRef]
50. Jeanquartier, F.; Jean-Quartier, C.; Stryeck, S.; Holzinger, A. Open Data to Support CANCER Science—A Bioinformatics Perspective on Glioma Research. *Onco* **2021**, *1*, 219–229. [CrossRef]
51. Bunda, S.; Heir, P.; Metcalf, J.; Li, A.S.C.; Agnihotri, S.; Pusch, S.; Yasin, M.; Li, M.; Burrell, K.; Mansouri, S.; et al. CIC protein instability contributes to tumorigenesis in glioblastoma. *Nat. Commun.* **2019**, *10*, 1–17. [CrossRef]
52. Appin, C.L.; Brat, D.J. Biomarker-driven diagnosis of diffuse gliomas. *Mol. Asp. Med.* **2015**, *45*, 87–96. [CrossRef]
53. Hu, W.; Duan, H.; Zhong, S.; Zeng, J.; Mou, Y. High Frequency of PDGFRA and MUC Family Gene Mutations in Diffuse Hemispheric Glioma, H3 G34-mutant: A Glimmer of Hope? 2021. Available online: https://assets.researchsquare.com/files/rs-904972/v1/2e19b03a-6ecb-49e0-9db8-da9aaa6d7f11.pdf?c=1636675718 (accessed on 12 January 2022).
54. Wong, W.H.; Junck, L.; Druley, T.E.; Gutmann, D.H. NF1 glioblastoma clonal profiling reveals KMT2B mutations as potential somatic oncogenic events. *Neurology* **2019**, *93*, 1067–1069. [CrossRef]
55. Hai, L.; Zhang, C.; Li, T.; Zhou, X.; Liu, B.; Li, S.; Zhu, M.; Lin, Y.; Yu, S.; Zhang, K.; et al. Notch1 is a prognostic factor that is distinctly activated in the classical and proneural subtype of glioblastoma and that promotes glioma cell survival via the NF-κB (p65) pathway. *Cell Death Dis.* **2018**, *9*, 1–13. [CrossRef]
56. Romo, C.G.; Palsgrove, D.N.; Sivakumar, A.; Elledge, C.R.; Kleinberg, L.R.; Chaichana, K.L.; Gocke, C.D.; Rodriguez, F.J.; Holdhoff, M. Widely metastatic IDH1-mutant glioblastoma with oligodendroglial features and atypical molecular findings: A case report and review of current challenges in molecular diagnostics. *Diagn. Pathol.* **2019**, *14*, 1–10. [CrossRef]
57. Haas, B.R.; Cuddapah, V.A.; Watkins, S.; Rohn, K.J.; Dy, T.E.; Sontheimer, H. With-No-Lysine Kinase 3 (WNK3) stimulates glioma invasion by regulating cell volume. *Am. J. Physiol. Cell Physiol.* **2011**, *301*, C1150–C1160. [CrossRef]
58. Suzuki, H.; Aoki, K.; Chiba, K.; Sato, Y.; Shiozawa, Y.; Shiraishi, Y.; Shimamura, T.; Niida, A.; Motomura, K.; Ohka, F.; et al. Mutational landscape and clonal architecture in grade II and III gliomas. *Nat. Genet.* **2015**, *47*, 458–468. [CrossRef]
59. Puustinen, P.; Keldsbo, A.; Corcelle-Termeau, E.; Ngoei, K.; Sønder, S.L.; Farkas, T.; Kaae Andersen, K.; Oakhill, J.S.; Jäättelä, M. DNA-dependent protein kinase regulates lysosomal AMP-dependent protein kinase activation and autophagy. *Autophagy* **2020**, *16*, 1871–1888. [CrossRef]
60. Stucklin, A.S.G.; Ryall, S.; Fukuoka, K.; Zapotocky, M.; Lassaletta, A.; Li, C.; Bridge, T.; Kim, B.; Arnoldo, A.; Kowalski, P.E.; et al. Alterations in ALK/ROS1/NTRK/MET drive a group of infantile hemispheric gliomas. *Nat. Commun.* **2019**, *10*, 1–13.
61. Franceschi, S.; Lessi, F.; Aretini, P.; Ortenzi, V.; Scatena, C.; Menicagli, M.; La Ferla, M.; Civita, P.; Zavaglia, K.; Scopelliti, C.; et al. Cancer astrocytes have a more conserved molecular status in long recurrence free survival (RFS) IDH1 wild-type glioblastoma patients: New emerging cancer players. *Oncotarget* **2018**, *9*, 24014. [CrossRef]
62. Wang, Y.; Wang, L.; Blümcke, I.; Zhang, W.; Fu, Y.; Shan, Y.; Piao, Y.; Zhao, G. Integrated genotype-phenotype analysis of long-term epilepsy-associated ganglioglioma. *Brain Pathol.* **2021**, *32*, e13011. [CrossRef]
63. Xiao, M.; Du, C.; Zhang, C.; Zhang, X.; Li, S.; Zhang, D.; Jia, W. Bioinformatics analysis of the prognostic value of NEK8 and its effects on immune cell infiltration in glioma. *J. Cell. Mol. Med.* **2021**, *25*, 8748–8763. [CrossRef]
64. Holzinger, A. Explainable ai and multi-modal causability in medicine. *i-com* **2020**, *19*, 171–179. [CrossRef]