Pathway-informed deep learning model for survival analysis and pathological classification of gliomas

Gun Kaynar[1, ‡] Doruk Cakmakci[2, ‡] Caroline Bund[3,4,5] Julien Todeschi[6]
Izzie Jacques Namer[3,4,5] A. Ercument Cicek[1,7,*]

[1]Computer Engineering Department, Bilkent University, Ankara, Turkey 06800
[2]School of Computer Science, McGill University, Montreal, QC, Canada
[3]MNMS Platform, University Hospitals of Strasbourg, France
[4]ICube, University of Strasbourg / CNRS UMR 7357, Strasbourg, France
[5]Department of Nuclear Medicine and Molecular Imaging, ICANS, Strasbourg, France
[6]Department of Neurosurgery, Strasbourg University Hospital, Strasbourg, France
[7]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213
[*]Correspondence: cicek@cs.bilkent.edu.tr, [‡]Equal contribution

Abstract

Online assessment of the tumor pathology during surgery is an important task to give feedback to the surgeon who can be more liberal or conservative in the resection based on the input. While there are methods that perform metabolomics-based online tumor grade prediction, their performance and model complexities are limited by the small dataset sizes. Here, we propose a pathway-informed deep learning model, PiDeeL, to perform survival analysis simultaneously for a better prognostic assessment. We show that incorporating the pathway information into the model architecture and using the multitask learning framework reduce the model complexity and enable us to use deeper architectures with better performance. We show that PiDeeL improves the grade prediction performance of the state-of-the-art in terms of the Area Under the ROC Curve (AUC-ROC) by 3.38% and the Area Under the Precision-Recall Curve (AUC-PR) by 4.06% and survival analysis performance based on the time-dependent concordance index (c-index) by 2.06%. Analyzing the importance of input metabolites and neurons representing pathways provides insights into tumor metabolism. We foresee that using this model in the surgery room will help surgeons adjust the surgery plan on the fly and will result in improved prognosis due to optimized surgical procedures. The code is released at `https://github.com/ciceklab/PiDeeL`. The data used in this study is released at `https://zenodo.org/record/7228791`.

1 Introduction

Gliomas have uncertain patient mortality rates due to differences in suspected brain cell types and high of levels variance in cancer progression dynamics [1]. 2021 version of central nervous system tumor taxonomy from the World Health Organization (WHO) [2] classifies gliomas according to detected cancerous cell types in the tumor tissue sample resected from the patient's brain (e.g. astrocytes, oligodendroglioma) as well as four pathologically different grades (i.e. grade I - IV). Furthermore, this taxonomy takes differential histopathologic, molecular, and genetic characteristics of tumors into account. The resulting tumor grading scheme inherently orders grades concerning increasing malignancy and decreasing prognostic optimism. After factoring in the high prevalence of gliomas among brain tumors diagnosed in adults [3], devising a surgical management strategy that aims to aid personalized clinical decision-making via an intraoperative or preoperative surgeon feedback mechanism becomes essential.

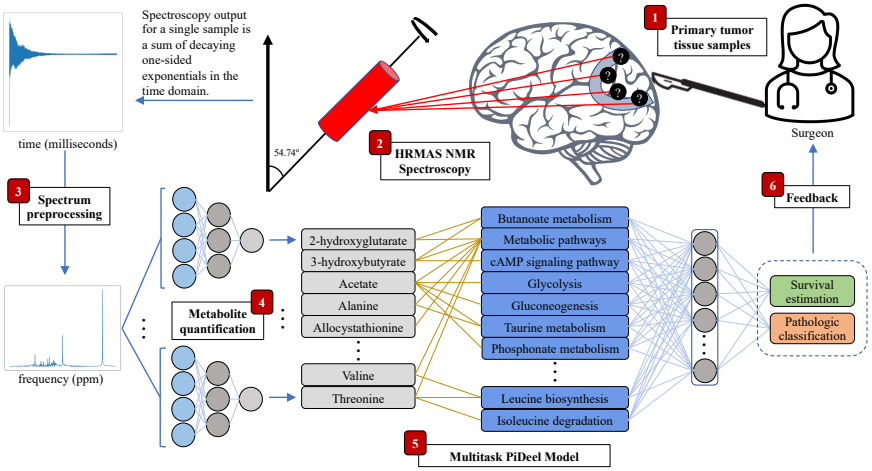

Figure 1: PiDeeL assisted intraoperative feedback loop. (Step 1) The surgeon first removes primary tumor tissue from the patient's brain as much as possible and prepares specimens from the resected tissue. (Step 2) Prepared samples are then sent for HRMAS NMR spectroscopy analysis without further processing. (Step 3) Time domain signals resulting from the spectroscopy of each sample are preprocessed to uncover metabolite peaks harbored in the corresponding frequency domain spectrum. (Step 4) 37 metabolites are quantified from preprocessed spectra in the order of milliseconds via metabolite-specific 2-layer fully-connected neural networks. (Step 5) PiDeeL transforms metabolite concentrations to survival and pathological classification predictions also in the order of milliseconds. (Step 6) Based on this feedback, the surgeon decides to preempt or continue performing the surgery, again with the guidance of PiDeeL.

To amplify the percentage of tumor tissue resected from a patient's brain, various imaging [4], spectroscopy as well as mass spectrometry [5–13], and optical spectrometry [14–26] based techniques, were proposed in the literature. As a member of this cohort, ^{1}H High Resolution Magic Angle Spinning Nuclear Magnetic Resonance (HRMAS NMR) spectroscopy measures metabolic activity from an unprocessed tissue specimen [27] with a turnaround time of $\sim$ 20 mins as a time domain signal (Free Induction Decay - FID).

After proper preprocessing steps, the HRMAS NMR spectrum becomes a metabolic information source tailored for deducing the presence of left-over tumor tissues in the excision cavity and providing feedback to the surgeon about the nature of the tumor. This intraoperative feedback loop can be established with the presence of a pathologist knowledgeable on tumor metabolism, and an NMR technician who detects an abundance of a few metabolites with the pathologist. However, the feedback resulting from this strategy is limited by (1) the availability of the experts during the surgery timeframe; and (2) the throughput of the metabolite quantification procedure manually performed by the technician one metabolite at a time for each tissue specimen. Moreover, overlapping peaks may not be fully separated in the preprocessed ^{1}H spectra [28].

Data-driven computational modeling (i.e. machine learning) has the promise to be a valuable companion to the manual feedback loop depicted above due to the high-dimensional, complex and structured nature of preprocessed HRMAS NMR spectra. Indeed, various supervised machine learning models have already been explored in this context to provide automated feedback with regard to pathological classification [29,30] of tissue specimens resected during glioma surgery. The dataset used for these studies [31] consists of HRMAS NMR spectra of 565 glioma and healthy

control tissue samples. It is relatively small despite being the largest of its kind which prohibited the use of complex models and forced the models built to be shallow. This suggests room for improvement in the performance of the feedback pipeline.

From a methodological perspective, the applicability of deep learning models in small data regimes can be improved by making informed architectural choices to simplify models. Incorporating biological prior information may limit the connectivity of the neurons in a model and simultaneously (1) alleviate overfitting when available dataset sizes are small, and (2) aid model interpretability regardless of the dataset size. While there is no application in the metabolomics domain, biologically informed deep learning architectures were successfully utilized in the context of metastasis prediction in patients with prostate cancer based on their genomic profiles [32], and chromatin accessibility prediction based on reference genome sequences while accounting for known transcription factor motifs during weight initialization step [33].

Furthermore, the pathological classification of tumors may not be the only determinant of multifaceted and patient-specific cancer progression semantics. In the case of glioblastoma multiforme (GBM), a grade IV astrocytic glioma type, a dismal prognosis is often expected due to the highly malignant and invasive nature of this class of tumors [1]. Despite their severity, prolonged and progression-free survival was observed in isocitrate dehydrogenase (IDH) wildtype GBM [34–36]. This finding promotes the inclusion of survival risk estimates in the feedback given to surgeons for more accurate surgical decision-making.

Here, we propose a deep neural network architecture to predict the survival of the glioma patient and the pathological classification of the tumor using HRMAS NMR signal to give feedback to the surgeons. While there were methods proposed for automized pathological classification, this is the first study to provide these two orthogonal sources of information at the same time. For the first time in the metabolomics domain, we incorporate prior biological information from the metabolic networks into the neural network architecture; and also use multitask learning to decrease the number of parameters to be learned in our deep model and increase the interpretability of the result which is also important for the surgeon to understand the underlying biology. We show that this sparse but deep model (PiDeeL) achieves better performance compared to the state-of-the-art model in pathological classification and also provides survival prediction with a c-index of 68.7%. We analyze the most important metabolites and pathways for the reasoning of the model. We release the code at `https://github.com/ciceklab/PiDeeL` and the dataset at `https://zenodo.org/record/7228791`.

2 Methods

2.1 Dataset

The data used in this study has been collected at the University Hospitals of Strasbourg. The details of data collection and ethics statement are shared in Section 1.1 of the Supplementary Text. The age, tumor subtype, and grade distribution and details are also available in Supplementary Figure 1. From the acquired ERETIC-CPMG spectra of 565 glioma and healthy control samples, we prepare a sample set by taking out the control samples and the samples without survival information (event and duration). Our sample set consists of 384 glioma samples which contain 299 (77.9%) aggressive and 85 (22.1%) benign tumors. Among the glioma patients in our sample set, 269 (70.1%) of them were reported deceased before the acquisition (until January 2021) of the dataset, whereas 115 (29.9%) of them were not. Note that, the ERETIC-CPMG HRMAS NMR FID spectra of 384 glioma samples are released at `https://zenodo.org/record/7228791` with their corresponding

sample IDs. To prepare our FID spectra for automated metabolite quantification, we apply the same preprocessing pipeline in [30]. The details of preprocessing method with the parameters used can be found in Section 1.1.3 of the Supplementary Text.

2.2 Problem Formulation

Please see the Section 1.3 of the Supplementary Text for the automated metabolite quantification. The feature vector for sample i, $\widehat{M}^{(i)}$, is a 37-dimensional vector $[\widehat{M}_1^{(i)}, \widehat{M}_2^{(i)}, \widehat{M}_3^{(i)}, ..., \widehat{M}_{37}^{(i)}] \in \mathbb{R}^{37}$ where $\widehat{M}_j^{(i)}$ corresponds to the predicted level for the jth metabolite. On the other hand, the class labels for sample i are the event variable, survival duration and tumor malignancy, denoted with $Y_e^{(i)}$, $Y_d^{(i)}$ and $Y_m^{(i)}$, respectively. $Y_e^{(i)}$, is set to 1 if patient (i) is deceased, and 0 otherwise. The class label duration, $Y_d^{(i)}$, is the time between the date of the tumor removal surgery and the date of decease or the last control, in days. The class label $Y_m^{(i)}$ is set to 1 if sample i is from an aggressive tumor, and 0 otherwise. We train survival analysis and pathological classification models based on this formulation. The feature vector we use is the same in the two tasks. However, for survival analysis, we use $Y_e^{(i)}$ and $Y_d^{(i)}$ as our class labels. On the other hand, for the pathological classification, we use $Y_m^{(i)}$ as the class label. Finally, for the multitask learning model, we use all class labels simultaneously (multi-label classification).

For the survival analysis task, we refer to Cox's baseline hazard function $\lambda(t, \widehat{M}^{(i)})$ for sample i in the dataset, where t is the time. We assume that the proportional hazards function from Cox's model is a product of two functions: a baseline hazard function, $h_0(t)$ (time-dependent), and a risk function, $h(\widehat{M}^{(i)})$ (sample-dependent). The relation between the proportional hazards function, baseline hazard function, and risk function is given below [37], [38]:

$$\lambda(t, \widehat{M}^{(i)}) = h_0(t).e^{h(\widehat{M}^{(i)})} \tag{1}$$

Our goal is to predict the risk scores of glioma patients, so that the models we train for this task approximate a function f such that, $f(\widehat{M}^{(1:N)}) = h(\widehat{M}^{(1:N)})$. Diversely, for the pathological classification task, our models approximate a function g such that, $g(\widehat{M}^{(1:N)}) = Y_m^{(1:N)}$.

2.3 Survival Analysis

2.3.1 Baseline Models

Cox Proportional Hazards Model (Cox-PH) is a parametric linear model that estimates $h(\widehat{M}^{(i)})$. The predicted risk function, $\widehat{h}_\beta(\widehat{M}^{(i)})$ with parameter vector β, is proposed by [39] as:

$$\widehat{h}_\beta(\widehat{M}^{(i)}) = \beta^t \widehat{M}^{(i)} \tag{2}$$

where $\widehat{h}_\beta(\widehat{M}^{(i)})$ approximates the risk of patient i at the time t by estimating the parameter vector β. The approximation routine involves maximizing Cox partial likelihood which is given below [37], [38], [39]:

$$L(\beta) = \prod_{i:Y_e^{(i)}=1} \frac{exp(\widehat{h}_\beta(\widehat{M}^{(i)}))}{\sum_{j:Y_d^{(i)}>t} exp(\widehat{h}_\beta(\widehat{M}^{(j)})))} \tag{3}$$

Further, the Random Survival Forest Model (RSF) is a nonlinear ensemble method to estimate the cumulative hazard function, $\Lambda(t)$, which is defined below [40]:

$$\Lambda(t) = \int_0^t \lambda(u)\, du \tag{4}$$

The neural network approach to survival analysis is first proposed by Faraggi-Simon [37], then deep models are made available by DeepSurv [38]. The proposed loss function of these articles is the negative log partial likelihood of Equation 3. We define the loss function below with our formulation:

$$l(\beta) = -\sum_{i:Y_e^{(i)}=1} (\widehat{h}_\beta(\widehat{M}^{(i)}) - log \sum_{j:Y_d^{(i)}>t} e^{(\widehat{h}_\beta(\widehat{M}^{(j)}))}) \tag{5}$$

We fit Cox-PH, RSF, and DeepSurv models to our data to get a baseline. We train 2, 3, and 4-layer DeepSurv models with 64 neurons in hidden layers and ReLU activation as the nonlinearity function. We have a single neuron in the last layer to output the predicted risk scores of samples.

The trained models on our dataset of size N can be summarized as follows:

$$X^{(1:N)} = ReLU(FC^{(64)}(\widehat{M}^{(1:N)})) \tag{6}$$

$$\widehat{h}^{(1:N)} = FC^{(1)}(X^{(1:N)}) \tag{7}$$

where X represents the output of hidden layers, $\widehat{h}^{(i)}$ represents the predicted risk score, $FC^{(\cdot)}$ represents a fully-connected layer with $(\cdot)$ neurons and $ReLU$ represents the rectified linear unit. This summary is for a 2-layered DeepSurv model. Note that step 6 is repeated when more layers are added.

2.3.2 Pathway-informed Deep Learning Model (PiDeeL)

We introduce our model, PiDeeL, which is a pathway-informed multi-layer perceptron specially designed to exploit the biomarker metabolite level information on the input layer. We refer to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to parse the metabolite-to-pathway mapping and construct a 37 by 138 matrix (named PI_matrix where the value in the ith row and j^{th} column of matrix PI_matrix is 1 if the metabolite i appears in the pathway j, and 0 otherwise. Our model architecture is as follows:

$$PL^{(138)} = \left[\ FC^{(138)} weights\ \right] \left[PI_matrix\right] \tag{8}$$

$$PA^{(1:N)} = ReLU(PL^{(138)}(\widehat{M}^{(1:N)}) \tag{9}$$

$$X^{(1:N)} = ReLU(FC^{(64)}(PA^{(1:N)})) \tag{10}$$

$$\widehat{h}^{(1:N)} = FC^{(1)}(X^{(1:N)}) \tag{11}$$

where $PL^{(138)}$ represents the weights of the pathway-informed layer, $PA^{(1:N)}$ represents the activations of the first hidden layer (we name it pathway-activation), X represents the output of hidden layers, $\widehat{h}^{(i)}$ represents the predicted risk score, $FC^{(\cdot)}$ represents a fully-connected layer with $(\cdot)$ neurons and $ReLU$ represents the rectified linear unit. Note that the architecture of the 3-layer PiDeeL is given here. Step 10 is canceled in the 2-layer model, whereas it is repeated in the 4-layer model. In step 8, the matrix product of the PI_matrix and weights of $FC^{(138)}$ is taken. Then we feed $\widehat{M}^{(1:N)}$ into the model. That is, the first hidden layer becomes sparse with only the targeted metabolite-to-pathway connections.

We also train DeepSurv models with 138 neurons in the first hidden layer to compare the dense and sparse structures in the model. Similarly, we construct randomly connected $PI-matrices$ and name them $pseudo-PI-matrices$, and train PiDeeL with them to compare targeted and random sparsity in our model. The number of connections in the sparse layer, i.e. the sum of all the binary values in the PI_matrix is 468. Note that, the total connections in the $pseudo-PI-matrices$ are also 468.

2.4 Pathological Classification

We train fully-connected models and PiDeeL models with weighted binary cross-entropy loss. The architecture of the models is the same as 2.3. However, in the output neuron, we use sigmoid activation.

2.5 Multitask Learning of PiDeeL for Survival Analysis and Pathological Classification

We train 2-layer 3-layer and 4-layer PiDeeL models with 2 neurons in the output layer. We use the loss function from Equation 5 in the first output neuron and weighted binary cross entropy loss in the second one. The model can be summarized as follows:

$$PL^{(138)} = \left[\ FC^{(138)} weights\ \right] \left[PI_matrix\right] \tag{12}$$

$$PA^{(1:N)} = ReLU(PL^{(138)}(\widehat{M}^{(1:N)}) \tag{13}$$

$$X^{(1:N)} = ReLU(FC^{(64)}(PA^{(1:N)})) \tag{14}$$

$$\widehat{h}^{(1:N)}, \widehat{Y}_m^{(1:N)} = FC^{(2)}(X^{(1:N)}) \tag{15}$$

where $PL^{(138)}$ represents the weights of the pathway-informed layer, $PA^{(1:N)}$ represents the activations of the first hidden layer, X represents the output of hidden layers, $\widehat{h}^{(i)}$ represents the predicted risk score, $\widehat{Y}_m^{(1:N)}$ represents the predicted tumor malignancy, $FC^{(\cdot)}$ represents a fully-connected layer with $(\cdot)$ neurons and $ReLU$ represents the rectified linear unit.

2.6 Metabolite and pathway importance analysis

For the interpretability of our models, we perform SHapley Additive exPlanations (SHAP) analysis on our feature vector $\widehat{M}^{(1:N)}$ for assessing the importance of each metabolite on the output of the multitask PiDeeL. We also calculate SHAP values of $PA^{(1:N)}$ for analyzing the importance of each pathway on the prediction of the multitask PiDeeL.

3 Results

3.1 Experimental Setup

In this section, we describe the experimental setup used for evaluating PiDeeL as well as baseline models formulated in Section 2.3.1 context of survival analysis and pathological classification scenarios. We measured the performance of survival analysis models in the unit of time-dependent concordance index (c-index) described in [41]. The original concordance index [42] is computed with the individually predicted patient risk scores, whereas with a time-dependent definition, individual risk scores are calculated with the predicted risk function and the baseline hazard function that is computed on the sample set. [41], [43] Furthermore, we assess performance pathological classification models with respect to Area Under the ROC Curve (AUC-ROC) and Area Under the Precision-Recall Curve (AUC-PR) metrics. For each task and model type, we adopt a 5-fold cross-validation routine and repeat it 3 times with different seeds (i.e. 15 iterations). Before each iteration, we shuffle and divide our dataset to training, validation and test splits such that no sample or patient are shared across folds. The results discussed hereafter are with respect to the predictive capability of models on the test split. During iteration i of the cross-validation setup, Cox-PH and RSF models were tested on the fold i, validated on fold $(i+1) mod 5$, and trained on the remaining 3 folds. We re-trained the selected configuration on the union of training and validation splits to get stronger baseline models. For all deep learning models, re-training steps were

not performed due to the presence of early-stopping dynamics. Please see Section 1.2.1, 1.2.2 and 1.2.3 of Supplementary Text for hyperparameter space for the models, hyperparameter selection, implementation details as well as hyper

3.2 Survival Analysis

Please see panel (a) of Figure 2 for the survival estimation performance of Cox-PH and RSF (gray background) as well as DeepSurv (green background) and PiDeeL (lavender background) with respect to the c-index. We observe that 4-layer PiDeeL either slightly or substantially outperforms all baselines and DeepSurv models on survival estimation and achieves a median c-index of 68.7%. We, then, see that performance minimally increases when the depth of PiDeeL is increased from 2 to 4 (improvement up to median c-index of 0.4%). On the other hand, the performance of DeepSurv models diminishes with increasing model depth (deterioration up to a median c-index of 1.5%). Furthermore, when DeepSurv and PiDeeL with the same number of layers are compared, we see the performance gap increase with respect to the number of layers in the model. Here, we quantitatively observe the benefit of incorporating biological information in modeling choices. For 2,3 and 4-layered networks, PiDeeL improves upon DeepSurv having the same model depth by 2.5%, 3.4%, and 4.3% according to the median c-index, respectively. This phenomenon supports, for survival estimation task, the robustness of sparse architectures to overfitting in small data regimes, which is often the case with clinical applications of metabolomics to understand tumor metabolism. When the performance of baseline models is considered in this setup, RSF appears to be a particularly strong baseline with a median c-index of 67.8%. This is also expected since the power of random forest-based models on HRMAS NMR metabolomics data has been demonstrated previously on a cohort of similar size [29, 30]. Still, PiDeeL achieves better performance than RSF while providing information about possibly active metabolic pathways by learning to pool information from metabolite sets encoded in its first layer. Finally, the Cox-PH model substantially performs worse than RSF and, in turn, PiDeeL. This may indicate a violation of the proportional hazards assumption implicitly made by the model.

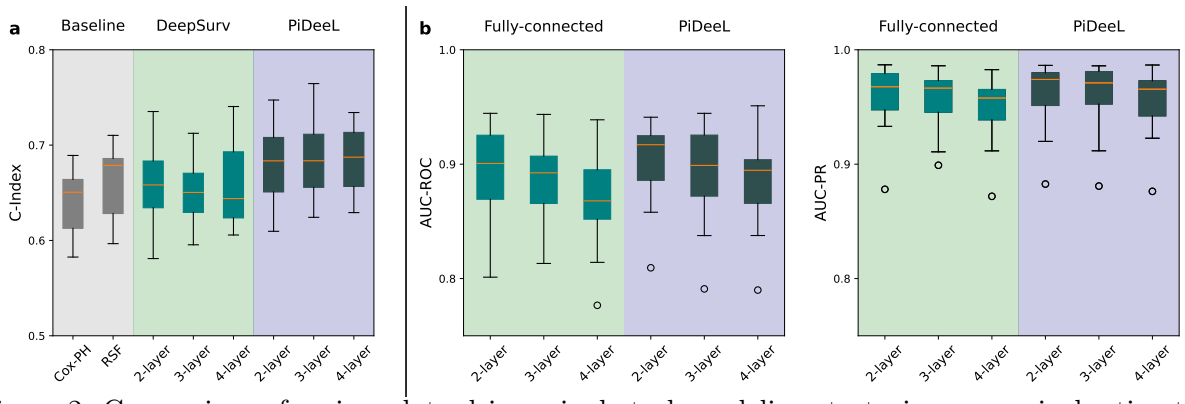

Figure 2: Comparison of various data-driven single-task modeling strategies on survival estimation in terms of c-index (Panel a) and pathologic classification in terms of AUC-ROC (Panel b - left) and AUC-PR (Panel b - right). This evaluation attempt is based on 5-fold cross-validation repeated 3 times.

3.3 Pathological Classification

Please see panel (b) of Figure 2 for a visual depiction of pathological classification performance attained by fully-connected networks (green background) and PiDeeL (lavender background) having 2, 3, or 4 layers. Here, similar to the survival analysis results presented in Section 3.2, we observe that PiDeeL consistently improves upon the fully-connected baseline, regardless of architecture depth or the performance metrics used for the current evaluation. In the case of AUC-ROC, we see 1.7%, 0.7% and 2.7% improvement in performance PiDeeL compared with fully-connected baseline having 2, 3, and 4 layers. As implicitly mentioned above, we see a similar trend for the AUC-PR metric where improvements are 0.6%, 0.4%, and 0.8% for the same order of model depth. However, for the current setup of the pathological classification task, we interestingly do not quantitatively observe PiDeeL's robustness to overfitting on small data regimes when the model deepens. In other words, when the number of layers in PiDeeL increases, performance on pathologic classification task worsens. Furthermore, we previously observed a similar phenomenon in DeepSurv models trained for survival analysis (Section 3.2). Note that it is meaningful for fully-connected networks discussed in this section to support this finding since they have the same architecture as DeepSurv except the output layer activation function. The discrepancy in PiDeeL's overfitting behavior between survival estimation and pathological classification tasks may indicate that the survival estimation task may have a more complex nature than pathological classification. This, in turn, reinforces the validity of the main takeaway from this work: Surgeon feedback may benefit from the presence of pathological classification and survival estimate at the same time.

3.4 Multi-task Learning of Survival Analysis and Pathological Classification via PiDeeL

Please see Figure 3 for the performance comparison of standalone (lavender background) and joint learning (red background) of survival estimation (Panel a) and pathological classification (Panel b) tasks using PiDeeL. Across all analyzed model depths, we observe that the multi-task modeling strategy is at least slightly beneficial for performance on both tasks at the same time. While 4-layer multitask PiDeeL attains the highest median c-index of 69.5% among others, 2-layer multitask PiDeeL is observed to outperform more complex alternatives on pathological classification with AUC-ROC of 91.9% and AUC-PR of 97.5%. This is in line with the task complexity and robustness to overfitting phenomena discussed in Section 3.3.

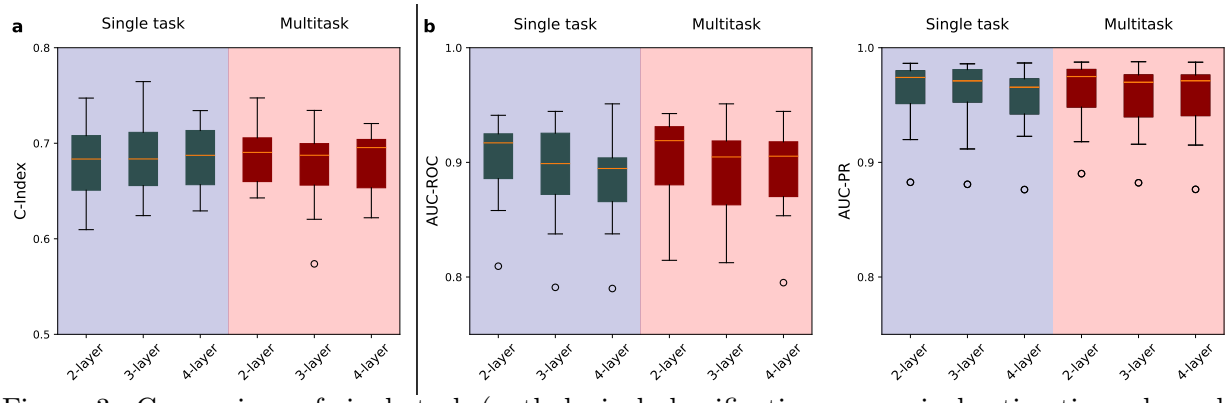

Figure 3: Comparison of single-task (pathological classification or survival estimation - lavender background) and multi-task (both tasks together - red background) of PiDeeL. This evaluation attempt is based on 5-fold cross-validation repeated 3 times.

3.5 Metabolite and Pathway Importance Analysis

When combined with posthoc feature importance calculation methods such as SHAP [44], the pathway-informed architecture of PiDeeL enables us to jointly analyze the importance of metabolites and pathways for the model output. In addition to quantitative analyses presented above, we perform feature importance analyses on multi-task 4-layer PiDeeL to empirically pinpoint metabolites and pathways candidate for being implicated in gliomas, their progression, and associated patient survival. We focus our interpretation attempts on the survival head of the model and calculate SHAP values on the input features (i.e. in silico predicted metabolite concentrations - $\widehat{M}^{(1:N)}$ in Eq. 1) as well as pathway-informed layer activations ($PA^{(1:N)}$ in Eq. 14).

Please see Supplementary Figure S3 for the metabolite importance plot for survival estimation head of multi-task PiDeeL. We highlight the top 3 metabolites having the highest mean SHAP value among all 37 metabolites according to our analyses in the decreasing order of importance *alanine*, *glutamine*, and *glutamate*. We also pinpoint the top 3 pathways having the highest SHAP values among the 138 pathways used for pathway-informed layer, in the decreasing order of importance as: *Mineral absorption*, *mTOR signaling pathway* and *Alanine— aspartate and glutamate metabolism*.

Through a case-control study, alanine was reported to be a potential biomarker for malignant gliomas [45]. Similarly, glutamate was also found to play a central role in malignant gliomas through multiple mechanisms [46]. Since malignant gliomas usually have a poor prognosis and are associated with dismal patient survival, the ranking of alanine and glutamate in the list is not surprising. Furthermore, glutamine was reported to serve as the primary precursor of glutamate, and to play a critical role in the normal functioning of the brain [47]. Glutamine metabolism can regulate multiple pathways including energy production and redox homeostasis in brain cancer cells [47]. Alterations in glutamine metabolism is suspected to be implicated in glioma metabolism and glutamine itself was coined as a candidate biomarker for glioma progression and response to treatment [48]. Thus, the rank of glutamine in the list is also expected.

Here, we biologically motivate the connection between the top 3 pathways and gliomas. Minerals and their absorption are fundamentally important to sustain life [49]. For example, calcium plays a wide variety of roles in the human body. Moreover, various natural dietary components were reported to have potential benefits in the prevention and treatment of gliomas [50]. Calcium was reported to be one of the dietary components having a potential for glioma prevention by affecting apoptosis and DNA repair [51]. Thus, the rank of *Mineral absorption* pathway according to calculated importance is biologically grounded. Furthermore, the mammalian target of rapamycin (mTOR), is reported to have a key role in integrating signal transduction and metabolic pathways in glioblastoma [52]. Since $\sim 46\%$ of the cohort analyzed in this work were diagnosed with glioblastoma, the importance rank of *mTOR signaling pathway* is viable. Finally, two of the three metabolites implicated in *Alanine— aspartate and glutamate metabolism* pathway, were also detected among the top 3 important metabolites with respect to metabolite-level importance scores, as discussed above. This qualitatively reinforces the predictive capabilities of PiDeeL and showcases its interpretability in both metabolite and pathway levels.

3.6 Ablation studies for pathway-informed architecture of PiDeeL

Besides the above-mentioned comparison of fully-connected networks and PiDeeL in various experimental settings, we performed two ablation studies on pathway-informed architecture trained for survival estimation to better see its benefits. First, we replaced 138 sparsely connected neurons in the first layer of PiDeeL with 138 densely connected neurons. Note that, the resulting architecture

is not pathway-informed and is very similar to the DeepSurv model discussed in Section 3.2 which had 64 densely connected neurons. Please see, panel a of Figure S2 comparison of these models. We see pathway-informed models outperform this version of pathway ablated model by $\sim 3.5\%$ across all model depths. Second, we randomly shuffled metabolite-pathway connections in the first layer of PiDeeL 150 times to obtain 150 different models. We train and test these models also on 5-fold cross-validation repeated 3 times. Please see, panel b of Figure S2 for the comparison of these models. Here, we also observe that PiDeeL outperforms the randomly ablated model consistently.

4 Discussion

1H HRMAS NMR spectrum has been a reliable resource for distinguishing malign and healthy samples taken from the excision cavity during surgery. It is a good fit because it is non-destructive and can analyze small samples of raw tissue specimens [27]. Machine learning techniques that learn from the NMR signal to distinguish healthy and tumor tissue, as well as benign and aggressive tumors, have been shown to be successful. Yet, the performance was limited due to small training set sizes [29, 30] which prohibit using complex architectures like deep neural networks which learn a hierarchical composition of complex features. We, for the first time, are able to use such a hierarchically complex model for distinguishing benign and aggressive tumors. Despite having a deep model, we are able to decrease the number of trainable parameters in the model by composing the architecture with prior biological information obtained from metabolic networks. We show better performance compared to the state-of-the-art result.

The pathology of the tumor is important for the surgeon who might want to resect more tumor tissue and risk the patient's well-being if the tumor is aggressive, or who might risk leaving some residual tumor tissue if the tumor is likely benign. Yet, this does not one-to-one correspond to survival. Thus, it is important to predict the survival of a patient and inform the surgeon about this as well. For this, we train the same model to predict the survival of patients with high performance. An example showing the usefulness of the survival prediction along with pathological classification is as follows: Our model is able to correctly predict long survival for a patient with aggressive glioma, who lived 5,656 days after surgery. The surgeon can be more conservative in this case compared to being informed with only a prediction of an aggressive tumor. Similarly, our model correctly predicts low survival for a patient with a benign tumor who lived only 43 days after surgery.

We also show that using a multitask learning architecture that basically combines two architectures and predicts both labels simultaneously further increases the performance for both tasks and helps the model to generalize better. As the two tasks are related, the model requires a relatively smaller number of samples to generalize better to unseen samples compared to individual single task models.

Our study is bottlenecked by the automated metabolite quantification model which has shown promising performance and generalizability. Yet, this component consists of individual multi-layered perceptrons for 37 metabolites for which a training dataset exists. We think our model would benefit from a more comprehensive spectrum of metabolites to in particular predict the survival of a patient which is a complex outcome to approximate. The human metabolome roughly consists of 3,000 metabolites and we are currently limited to using only 1% of this information source.

Acknowledgements

This work was supported by grants from BPI France (ExtempoRMN Project), Hôpitaux Universitaires de Strasbourg, Bruker BioSpin, Univ. de Strasbourg and the Centre National de la Recherche Scientifique; also by TUBA GEBIP, Bilim Akademisi BAGEP and TUSEB Research Incentive awards to AEC.

References

[1] Omuro, A. & DeAngelis, L. M. Glioblastoma and other malignant gliomas: a clinical review. *Jama* **310**, 1842–1850 (2013).

[2] Osborn, A., Louis, D., Poussaint, T., Linscott, L. & Salzman, K. The 2021 world health organization classification of tumors of the central nervous system: What neuroradiologists need to know. *American Journal of Neuroradiology* **43**, 928–937 (2022).

[3] Ostrom, Q. T. *et al.* Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2010–2014. *Neuro-oncology* **19**, v1–v88 (2017).

[4] Stummer, W., van den Bent, M. J. & Westphal, M. Cytoreductive surgery of glioblastoma as the key to successful adjuvant therapies: new arguments in an old discussion. *Acta neurochirurgica* **153**, 1211–1218 (2011).

[5] Brown, M. V. *et al.* Cancer detection and biopsy classification using concurrent histopathological and metabolomic analysis of core biopsies. *Genome medicine* **4**, 33 (2012).

[6] Schafer, K.-C. *et al.* Real time analysis of brain tissue by direct combination of ultrasonic surgical aspiration and sonic spray mass spectrometry. *Analytical Chemistry* **83**, 7729–7735 (2011).

[7] Eberlin, L. S. *et al.* Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. *Proceedings of the National Academy of Sciences* **110**, 1611–1616 (2013).

[8] Pirro, V. *et al.* Intraoperative assessment of tumor margins during glioma resection by desorption electrospray ionization-mass spectrometry. *Proceedings of the National Academy of Sciences* **114**, 6700–6705 (2017).

[9] Santagata, S. *et al.* Intraoperative mass spectrometry mapping of an onco-metabolite to guide brain tumor surgery. *Proceedings of the National Academy of Sciences* **111**, 11121–11126 (2014).

[10] Calligaris, D. *et al.* Molecular typing of meningiomas by desorption electrospray ionization mass spectrometry imaging for surgical decision-making. *International journal of mass spectrometry* **377**, 690–698 (2015).

[11] Calligaris, D. *et al.* Maldi mass spectrometry imaging analysis of pituitary adenomas for near-real-time tumor delineation. *Proceedings of the National Academy of Sciences* **112**, 9978–9983 (2015).

[12] Jarmusch, A. K. *et al.* Lipid and metabolite profiles of human brain tumors by desorption electrospray ionization-ms. *Proceedings of the National Academy of Sciences* **113**, 1486–1491 (2016).

[13] Fatou, B. *et al.* In vivo real-time mass spectrometry for guided surgery application. *Scientific reports* **6**, 1–14 (2016).

[14] Stummer, W. *et al.* Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase iii trial. *The lancet oncology* **7**, 392–401 (2006).

[15] Tsugu, A. *et al.* Impact of the combination of 5-aminolevulinic acid–induced fluorescence with intraoperative magnetic resonance imaging–guided surgery for glioma. *World neurosurgery* **76**, 120–127 (2011).

[16] Colditz, M. J. & Jeffree, R. L. Aminolevulinic acid (ala)–protoporphyrin ix fluorescence guided tumour resection. part 1: Clinical, radiological and pathological studies. *Journal of Clinical Neuroscience* **19**, 1471–1474 (2012).

[17] Montcel, B., Mahieu-Williame, L., Armoiry, X., Meyronet, D. & Guyotat, J. Two-peaked 5-ala-induced ppix fluorescence emission spectrum distinguishes glioblastomas from low grade gliomas and infiltrative component of glioblastomas. *Biomedical optics express* **4**, 548–558 (2013).

[18] Li, Y., Rey-Dios, R., Roberts, D. W., Valdés, P. A. & Cohen-Gadol, A. A. Intraoperative fluorescence-guided resection of high-grade gliomas: a comparison of the present techniques and evolution of future strategies. *World neurosurgery* **82**, 175–185 (2014).

[19] Lu, F.-K. *et al.* Label-free neurosurgical pathology with stimulated raman imaging. *Cancer research* **76**, 3451–3462 (2016).

[20] Jermyn, M. *et al.* Highly accurate detection of cancer in situ with intraoperative, label-free, multimodal optical spectroscopy. *Cancer research* **77**, 3942–3950 (2017).

[21] Jermyn, M. *et al.* Raman spectroscopy detects distant invasive brain cancer cells centimeters beyond mri capability in humans. *Biomedical optics express* **7**, 5129–5137 (2016).

[22] Chan, D. T. M., Sonia, H. Y.-P. & Poon, W. S. 5-aminolevulinic acid fluorescence guided resection of malignant glioma: Hong kong experience. *Asian journal of surgery* **41**, 467–472 (2018).

[23] Orringer, D. A. *et al.* Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated raman scattering microscopy. *Nature biomedical engineering* **1**, 0027 (2017).

[24] Poulon, F. *et al.* Optical properties, spectral, and lifetime measurements of central nervous system tumors in humans. *Scientific reports* **7**, 1–8 (2017).

[25] Hollon, T. C. *et al.* Rapid intraoperative diagnosis of pediatric brain tumors using stimulated raman histology. *Cancer research* **78**, 278–289 (2018).

[26] Xue, Z. *et al.* Fluorescein-guided surgery for pediatric brainstem gliomas: Preliminary study and technical notes. *Journal of Neurological Surgery Part B: Skull Base* **79**, S340–S346 (2018).

[27] Gogiashvili, M. *et al.* Hr-mas nmr based quantitative metabolomics in breast cancer. *Metabolites* **9**, 19 (2019).

[28] Karakaslar, E. O., Coskun, B., Outilaft, H., Namer, I. J. & Cicek, A. E. Predicting carbon spectrum in heteronuclear single quantum coherence spectroscopy for online feedback during surgery. *IEEE/ACM transactions on computational biology and bioinformatics* **17**, 719–725 (2019).

[29] Cakmakci, D. *et al.* Machine learning assisted intraoperative assessment of brain tumor margins using hrmas nmr spectroscopy. *PLoS Computational Biology* **16**, e1008184 (2020).

[30] Cakmakci, D. *et al.* Targeted metabolomics analyses for brain tumor margin assessment during surgery. *Bioinformatics* (2022).

[31] Cakmakci, D. *et al.* 1h hrmas nmr dataset of glioma and control samples (2020). URL `https://doi.org/10.5281/zenodo.3951448`.

[32] Elmarakeby, H. A. *et al.* Biologically informed deep neural network for prostate cancer discovery. *Nature* **598**, 348–352 (2021).

[33] Novakovsky, G., Dexter, N., Libbrecht, M. W., Wasserman, W. W. & Mostafavi, S. Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nature Reviews Genetics* 1–13 (2022).

[34] Poon, M. T., Sudlow, C. L., Figueroa, J. D. & Brennan, P. M. Longer-term (2 years) survival in patients with glioblastoma in population-based studies pre-and post-2005: a systematic review and meta-analysis. *Scientific reports* **10**, 1–10 (2020).

[35] Burgenske, D. M. *et al.* Molecular profiling of long-term idh-wildtype glioblastoma survivors. *Neuro-oncology* **21**, 1458–1469 (2019).

[36] Wong, D. & Yip, S. Finding a four-leaf clover—identifying long-term survivors in idh-wildtype glioblastoma. *Neuro-oncology* **21**, 1352 (2019).

[37] Faraggi, D. & Simon, R. A neural network model for survival data. *Statistics in medicine* **14**, 73–82 (1995).

[38] Katzman, J. L. *et al.* Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology* **18**, 1–12 (2018).

[39] Cox, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **34**, 187–202 (1972).

[40] Ishwaran, H., Kogalur, U. B., Blackstone, E. H. & Lauer, M. S. Random survival forests. *The annals of applied statistics* **2**, 841–860 (2008).

[41] Antolini, L., Boracchi, P. & Biganzoli, E. A time-dependent discrimination index for survival data. *Statistics in medicine* **24**, 3927–3944 (2005).

[42] Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the yield of medical tests. *Jama* **247**, 2543–2546 (1982).

[43] Gerds, T. A., Kattan, M. W., Schumacher, M. & Yu, C. Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. *Statistics in medicine* **32**, 2173–2184 (2013).

[44] Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In Guyon, I. *et al.* (eds.) *Advances in Neural Information Processing Systems 30*, 4765–4774 (Curran Associates, Inc., 2017). URL `http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf`.

[45] Firdous, S. *et al.* Dysregulated alanine as a potential predictive marker of glioma—an insight from untargeted hrmas-nmr and machine learning data. *Metabolites* **11**, 507 (2021).

[46] de Groot, J. & Sontheimer, H. Glutamate and the biology of gliomas. *Glia* **59**, 1181–1189 (2011).

[47] Natarajan, S. K. & Venneti, S. Glutamine metabolism in brain tumors. *Cancers* **11**, 1628 (2019).

[48] Ekici, S., Risk, B. B., Neill, S. G., Shu, H.-K. & Fleischer, C. C. Characterization of dysregulated glutamine metabolism in human glioma tissue with 1h nmr. *Scientific reports* **10**, 1–10 (2020).

[49] Kegg pathway: Mineral absorption - reference pathway (ko) (usp.br). `https://www.genome.jp/pathway/hsa04978`. Accessed: 2022-10-16.

[50] Kyritsis, A. P., Bondy, M. L. & Levin, V. A. Modulation of glioma risk and progression by dietary nutrients and antiinflammatory agents. *Nutrition and cancer* **63**, 174–184 (2011).

[51] Tedeschi-Blok, N., Schwartzbaum, J., Lee, M., Miike, R. & Wrensch, M. Dietary calcium consumption and astrocytic glioma: the san francisco bay area adult glioma study, 1991-1995. *Nutrition and cancer* **39**, 196–203 (2001).

[52] Akhavan, D., Cloughesy, T. F. & Mischel, P. S. mtor signaling in glioblastoma: lessons learned from bench to bedside. *Neuro-oncology* **12**, 882–889 (2010).