# OVERALL SURVIVAL PREDICTION OF GLIOMA PATIENTS USING GENOMICS

M. R. Navodini Wijethilake

208012F



Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science by Research



Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka



September 2021

# DECLARATION

I, Navodini Wijethilake, declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature ***UOM Verified Signature*** Date: 23/09/2021

The above candidate has carried out research for the Masters thesis Dissertation under my supervision.

Name of Supervisor: Dr. Dulani Meedeniya

Signature of the Supervisor: Date: 23/09/2021

The above candidate has carried out research for the Masters thesis/Dissertation under my supervision.

Name of Supervisor: Dr. Charith Chithraranjan

Signature of the Supervisor: ***UOM Verified Signature*** 21

# ABSTRACT

**Overall survival prediction** is a vital task that will lead for better patient management in clinical practise. Existing approaches mainly focus on imaging based survival prediction, which is non invasive, and thus, easier to be implemented at the initial diagnosis stages. However, the advancements in the DNA/RNA technologies has given access to genomic and transcriptomic profiles of the gliomas, that directly reflect the molecular level alterations. Thus, in this work we mainly focus on using transcriptomic profiles for survival prediction, an area that has not been widely analysed yet for survival prediction. We utilize the gene expression and mutation profiles, while augmenting the recent Artificial Intelligence approaches, such as deep probabilistic programming and multi task learning for prognosis prediction. Thereby we do not just focus on the application, we also contribute with novel learning paradigms to improve the classification task performances. Nonetheless, we also focus on proposing a novel loss function, since architectural wise the state of art performance has been achieved for classification tasks.

In addition, we also investigate ability to employ radiomics, for subtype classification, that is also associated with survival. Since subtypes mainly rely on the genomic alterations, we found it useful to focus on imaging features ability to predict prognosis of glioma.

**Keywords**: Survival Prediction; Prognosis; Deep Learning; Multi-task Learning; Deep Probabilistic Programming

# ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisors Dr. Dulani Meedeniya and Dr. Charith Chithraranjan for the immense guidance provided to successfully finish this research. I'm extremely thankful for your patience, motivation, and immense knowledge given throughout the period. I firmly believe without your courageous support this would not be able to reach this stage.

I wish to thank Dr. Indika Perera for his valuable insights and guidance from the very beginning of this research. I would like to convey my gratitude to the entire staff of the Department of Computer Science and Engineering, both academic and non-academic for all their support given throughout the entire Masters course period. This research was supported by the University of Moratuwa Senate Research Grant. I would like to acknowledge the grant and other relevant parties who work hard to provide the required facilities to up bring the research facilities in Sri Lanka more specifically, by providing the financial support.

Nonetheless, I would like to thank my family for all the love and support.

**Thank you!**

# LIST OF ABBREVIATIONS

| | |
|---|---|
| GBM | Glioblastoma |
| DL | Deep Learning |
| ML | Machine Learning |
| IR | Importance Ranking |
| PCA | Principle Component Analysis |
| RFE | Recursive Feature Elimination |
| CC | Correlation Coefficient |
| RF | Random Forest |
| LR | Linear Regression |
| SVM | Support Vector Machine |
| XGB | eXtreme Gradient Boosting |
| ST | Survival Tree |
| ANN | Artificial Neural Network |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| KM | Keplen Meier |
| TCGA | The Cancer Genome Atlas |
| CNN | Convolutional Neural Network |
| CGGA | Chinese Glioma Genome Atlas |
| OBTS | the Ohio Brain Tumor Study |
| GEO | Gene ExpressionOmnibus |
| CPTAC | Clinical Proteomic Tumor Analysis Consortium |
| TCIA | TheCancer Imaging Archive |
| LGG | Lower Grade Glioma |
| CoxPH | Cox Proportional Hazard model |
| CE | Cross-Entropy |
| FL | Focal Loss |
| LSCE | Label Smoothing Cross-Entropy |
| LSF | Label Smoothing Focal |

| | |
|---|---|
| normLSF | Normalized Label Smoothing Focal |
| Acc. | Accuracy |
| Prec. | Precision |
| Sens. | Sensitivity |
| MTL | Multi Task Learning |

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# Chapter 1

# INTRODUCTION

## 1.1 Domain Overview

Glioma is a type of tumor that occurs in glial cells and other progenitor cells. Out of all primary brain and central nervous system tumors, gliomas represent 26.7%. Gliomas mostly arise in the brain, more specifically in the parietal, frontal and temporal lobes, and less likely in the remaining major lobe. occipital. Studies say that these gliomas can also occur in spinal code and spinal nerves known as cauda equina cerebrum [1]. Glioblastomas (GBM) is the most lethal histological type with 54.7 percentage of primary brain & central nervous systems gliomas are GBMs. GBMs & Astrocytomas, another common glioma histological type, combinely count for 75 percentage of gliomas. With increasing age, the occurrence ratio of gliomas reduces for the adolescents (young adults aged 15 to 19 years) and children. Roughly, 46.5 percentage of brain tumors belongs to glioma category for these age groups [2]. Anyway, the incidence rate gradually grows in the cases over 20 years with the highest number of incidences are reported in the age group of 85+ years. Moreover, median age is calculated as 65 years at the diagnosis of GBM patients. Nonetheless, GBMs, astrocytomas, & oligodendrogliomas occur in male population than the females according to the studies [3].

### 1.1.1 Classification based on underlying histology

Gliomas are initially classified based on the underlying histology, i.e. the morphological aspects in certain cell types of the effected region. The brain cells consists of an rigorous network that help maintaining the normal functioning in the brain [4].

- Astrocytes: The cells, which can be seen in the brain are the main connective tissue cells. In astrocytomas, the astrocytes in the specific region are

faced for similar morphological changes.

- Oligodendrocytes: The cells, also visible in the brain, cover the neuronal axons with a myelin sheath. Oligodendrogliomas develop in these.

- Ependymal cells: Ependydoma, a less frequently occuring tumor in humans, arise in these cells.

The astrocytomas, oligodendrogliomas, oligoastrocytomas, and ependymal, the different histological types of glioma were allocated into 4 grades by World Health Organization (WHO) considering the natural disease cause, anaplastic features and malignancy of the tumore [5].

- Grade I WHO gliomas: Minimal malignancy and slowly growing cancer. Pilocytic Astrocytomas are included into this category.

- Grade II WHO gliomas: Grows slowly, but sometimes, an invasive spreading can be seen in the brain. diffuse astrocytoma is a cancer that comes under WHO grade II.

- Grade III WHO gliomas: Quickly expanding gliomas, with histological attributes similar to anaplasia.

- Grade IV WHO gliomas: The maximal lethal, maligance type of glioma, mostly GBMs belong to this category and can be identified by the necrosis region and microvascular proliferation present in the tumor.



### 1.1.2 Classification based on molecular pathogenesis

Molecular pathogenesis is studied to identify the genomic and transcriptomic level changes or alterations in the tumor regions, that can lead for gliomas. This is slightly different from the histological categorizations and indications.

**Grade II and III (WHO) gliomas:** Grade II and III (WHO) gliomas, both together is combined as Lower Grade Gliomas (LGG). Although, Gliomas are separated into astrocytomas, oligodendrogliomas, and mixed oligoastrocytomas,

traditionally based on the histology, these 3 histological types can be grouped into 2 subtypes considering their molecular level alterations.

- Mutations occur in both tumor protein p53 (TP53) gene and $\alpha$- thalassemia/ mental retardation syndrome X-linked (ATRX) gene: the 'astrocytic' genotype.
- Co-deletion of chromosomal arms 1p and 19q and mutations occur in telomerase reverse transcriptase (TERT) promoter: the 'oligodendroglial' genotype.

Thus, Oligoastrocytomas histological category is not needed anymore [6]. Both the above mentioned genotypes depicts Isocitrate dehydrogenase 1 (IDH1) mutation commonly and Isocitrate dehydrogenase 2 (IDH2) mutation less frequently. Thus, the latest grouping of gliomas is into three categories considering the end result, the response for the treatments and the natural histories [4].

- IDH1 mutant, 1p/19q co-deleted tumors have the oligodendroglial histology: these tumors have a longer survival, i.e. a better prognosis.
- IDH1 mutant, 1p/19q non co-deleted tumors have mostly an astrocytic morphology: Intermediate level survival.
- IDH wild-type, 1p/19q non co-deleted tumors: Shorter survival, i.e. poor prognosis.

### 1.1.3 Mortality of Glioma patients

Gliomas, have different levels of mortality and morbidity depending on their histology, localization, and growth. The most malignant glioma, GBM has a 15 month median survival, despite the surgical resection and radiation therapy and chemotherapy followed [7]. Other than that, after 5 years from the initial diagnosis less than 3% of GBM patients survived [8]. In addition, WHO grade II and III gliomas have a 78.1 and 37.6 months median survival respectively [9]. The median survival based on the histological types, astrocytomas and oligodendrogliomas are 5.2 and 7.2 years, respectively [10].

### 1.1.4 Overall Glioma survival and Glioma survival

Overall survival of glioma patients is measured either from the date of the initial diagnosis of glioma or from the date that the treatments begin. The length of the time is calculated until the glioma patient is alive. This Overall Survival measurement is commonly considered when deciding the impact of a particular treatment given at a certain time point to the patient. The clinical protocols are arranged by the clinicians to maximize the overall survival always without trading off the quality of life of glioma patients. This quality of life of a glioma patient depends on couple of factors, such as the size and the location of the tumor, complications that arise with the surgical interventions followed, impact of radiotherapy & pharmacotherapy, and other Psychosocial matters, which can ultimately effect the survival of patients [4].

## 1.2 Overview of the Problem

Traditionally, the survival of glioma patients are predicted in the research context, by using the medical images such as Magnetic Resonance Images (MRI), Computed Tomography (CT), or Positron Emission Tomography (PET). These images are initially acquired and preprocessed where the feature extraction, and classification into short, medium and long survival groups is followed. At the beginning a semi automated approaches were common where segmentation and the features such as volume, texture and shape, extraction were performed manually and thereafter machine learning (ML)-based classifications were performed for survival prediction. State-of-art approaches follow fully automated deep learning based segmentation, feature extraction and ML based classification and regression prediction models development, where the maximum accuracy reported is around 70%. However, these imaging modalities are still unable to grab the intra-tumor heterogeneity, that is present at the molecular levels; Although, imaging approaches has illuminated the overall survival analysis research through its non-invasive behaviour. Also for gliomas, statistical analysis and risk estimation methods are introduced to estimate the risk and the probability of survival after

a certain time period. The risk is given as a value that uses the expression levels of factors, such as gene expression, related to survival. The survival can be even measured using nomograms. Yet, these approaches are not able to accurately predict the survival of glioma patients, in spite of the publications that come up with various parameters.

## 1.3 Motivation

With the new improvements and the development of the high throughput technologies such as microarray and DNA/RNA sequencing, the possibility of acquiring heterogeneous genomic features is possible and therefore, more precise survival analysis can be performed, overcoming the limitations in the traditional radiomic approaches [11]. Nonetheless, the advancements in Artificial Intelligence, where the capability of handling high dimensional data and models have been advanced, the field of patient management related to cancer can be enlightened [12]. This can lead to an absolute orientation in glioma prognosis estimation research in the future. Thus, we were motivated to explore the ability of utilizing various data types associated with glioma survival to develop different tools and algorithms to predict survival. Nonetheless, current medical AI related research is not just focusing the application. Mostly, a novel algorithms are necessary while proving the application level improvements. Thus, we were also motivated to develop in algorithms while proposing new directions for prognosis estimation of glioma patients.

Ultimately, we identify the limitations and challenges that might direct the state of art survival prediction approaches for a clinically plausible era.

## 1.4 Importance of this research

Quality of life of a glioma patient relies on several facts, such as the location and size of the tumor, complications that occur due to surgical interventions, effects of radiotherapy and pharmacotherapy, and Psychosocial consequences. These facts are associated with impaired neurocognitive functions, communication abilities,

and acute toxicities and functional deficits. Long term surviving patients, treated with radiotherapy, have a high risk for neurocognitive impairments. Moreover, standard treatments such as anti-seizure agents and corticosteroids, used in routine medications, cause nausea, fatigueness, drowsiness, and many more short term side effects. Meanwhile, they can also induce long term effects such as diabetes mellitus, depression, hypertension, and osteoporosis.

In addition, other than physical consequences, non-physical impairments also arise in spiritual, psychological, behavioral domains. Nevertheless, the families also get affected by the illness and side effects causing vulnerable psychological trauma and financial consequences. Therefore, it is essential for the clinicians, neurosurgeons, and oncologists to decide what treatments the patients require for a long term survival with minimum side effects. Sometimes, the patients will weaken due to side effects than to the effects occur due to the glioma. More often, clinicians estimate the survival of patients by clinical factors assessment, imaging assessment, and their experience. Researches show that these decisions are too much biased, inaccurate, and optimistic. This will also impact the patient and also the families to deal with the situation and prepare for the future outcome. Further, the resources such as medications, therapies are required to be allocated among the patients based on the requirement giving priority to the prognosis of each patient. Thus, accurate survival prediction is a predominant factor for better treatment planning and clinical management with optimal utilization of resources.

## 1.5 Objectives

- Develop a prognostic risk score model to estimate the risk of Glioma.
- Design models to predict overall survival with different genomic data.
- Develop a robust genomic based platform for survival prediction of glioma patients.
- Analyze the prognostic gene markers significant for the model behavior

utilizing explainable AI.

## 1.6 Contribution

Survival Prediction with gene expression data:

- Identify prognostic genes with Iterative Bayesian Model Averaging algorithm and propose a risk score prognosis model.
- Perform Survival Prediction using traditional Machine Learning tools, with the selected genes.
- Propose a Novel Bayesian Neural Network using Pyro for survival class prediction.

Survival prediction with mutation data:

- Propose a multitask learning deep learning model for glioma grade and survival class prediction.
- Optimize the MTL model while proposing the novel 'GradNorm Asynchronous Task-Aware Optimization' Algorithm.

Additionally,

- Identify radiomics associated with molecular subtypes of Glioma.
- Find the correlations between radiomics, genomics and survival.
- Predict subtypes using ML with radiomics extracted from Glioma images.



## 1.7 Publications

- **Wijethilake, N.**, Islam, M., Meedeniya, D., Chitraranjan, C., Perera, I., & Ren, H. (2020). Radiogenomics of glioblastoma: Identification of radiomics associated with molecular subtypes. In Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology (pp. 229-239). Springer, Cham.

- **Wijethilake, N.**, Meedeniya, D., Chitraranjan, C., & Perera, I. (2020, October). Survival prediction and risk estimation of Glioma patients using mRNA expressions. In 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE) (pp. 35-42). IEEE.

- **N. Wijethilake**, D. Meedeniya, C. Chitraranjan, I. Perera, M. Islam and H. Ren, "Glioma Survival Analysis Empowered With Data Engineering—A Survey," in IEEE Access, vol. 9, pp. 43168-43191, 2021, doi: 10.1109/ACCESS.2021.3065965.

# Chapter 2

# LITERATURE REVIEW

## 2.1 Data types used in Glioma survival analysis

### 2.1.1 Glioma Screening and Data Collection in clinical practice

There are multiple tests that are being used for diagnosis and screening purposes of brain tumors. Firstly, physical status of the patient is closely examined by the physicians, whereas they thoroughly go through the patient's medical history reports, patient's life style and family history. This is followed by the neurological tests, where the patients is questioned about the sensitivity of the central nervous system and associated effects that might arise with the tumor [13]. Most of the times, the eye sight of the patient gets effected due to the compression of the eye muscles with the progression of the tumor. Thus, to determine whether such disrupts were occurred, visual field tests are also performed on the patient in the initial diagnosis steps.

Thereafter imaging tests, such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Computed Tomography Scan (CT Scans), Angiograms and Single Photon Emission Tomography Scan (SPECT) are performed on the patient. Nonetheless, urine and blood tests are also taken to identify whether there any changes present in those media, that occur with the tumor [13].

Yet, whether the tumor is cancerous or not can be determined only by closely investigating or looking into the tissue samples of the tumor using a microscope. Normally these tissue samples are acquired either by surgical resection or using biopsy. The surgical resection is performed by surgeons mostly while closely watching the images of the tumor obtained from different imaging methods. In this the completely tumor is removed unlike obtaining a small part of the tumor when performing biopsy [14]. Pathologist, the experts who can identify the deformities that occur in the cell level, help the other clinicians such as oncologists

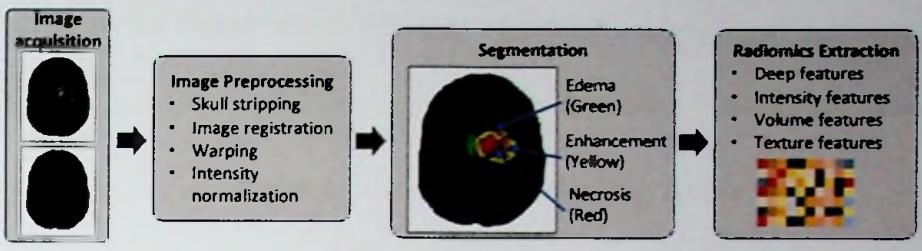



Figure 2.1: A general radiomics extraction pipeline.

to make decisions on the cancerous status of the patients through the obtained tissues from the tumor. In addition, the genomic profiles such as gene expression, mutation and methylation profiles are also acquired using the tissue samples of the tumor. We discuss about these platforms further in the section 2.1.3.

Besides there are various other tests performed throughout the diagnosing, screening, pre and post surgical stages of the glioma. These records are collected by many institutes for different purposes such as research. There are such publicly available data cohorts with patient records entirely for research purposes. These profiles of glioma patients are used in various studies for different tools and algorithm development, in order to aid the decision making process of the precision treatment planning and disease management.

### 2.1.2 Radiomics

During the initial diagnosis and screening of gliomas, various types of images are obtained to evaluate the size, location and the progression of the tumor. For survival related studies, these images are used. However, these raw images are pre processed before extracting features for such analysis. Next we are going to discuss the steps followed in the process of feature extraction. A generalized flow of these tasks are given in the Figure 2.1. There are 4 main steps in this process.

1. Acquisition of the images such as MRI.

2. Preprocessing of the images, where they are skull stripped, registered and intensity normalized.

3. Segment the tumor region into sub regions.

4. Feature extraction: Radiomics, i.e. imaging related features are obtained.

These 4 parts are discussed further in the following sections.

### MRI Acquisition

Traditionally survival prediction is done using the radiology images such as MRI. The non invasive approach followed in imaging methods, is an advantage in this radiomic based survival prediction. Thus, in cancer diagnosis and clinical monitoring protocols images are widely utilized to visualize the surface structure (morphology) and functional behavior of the tumor and the surrounding area. Out of different medical imaging tools, Magnetic Resonance Imaging (MRI) is the most famous radio-logical tool used in tumor visualization. There are 4 main types of modalities that are been frequently obtained. They are T1 weighted (T1), T2 weighted (T2), Fluid-attenuated inversion recovery (FLAIR), and gadolinium-enhanced (T1 contrast) sequences. More specifically each one of these elaborate different regions of the tumor. For an example the pathologically enhanced region is clearly visible in the Flair and T1 contrast image modalities [15]. These MRI sequences are collected along with the survival and other clinical information by many institutes. Some are publicly available for the research purposes while some are kept for their own research studies. We discuss the publicly available datasets further in section 2.1.5.

### MRI Preprocessing

The initially obtained MRI images consists of background non brain tissues, eye balls and skin. These are obstacles for obtaining precise segmentation of the tumor. Therefore, these unnecessary parts are removed first using skull stripping tools. Morphology based skull stripping methods, intensity and deformable intensity based skull stripping methods, hybrid approaches and atlas based skull stripping methods are few of those tools [16]. These approaches are embedded in

Table 2.1: Open source Tools available for Preprocessing and Segmentation of MRI

| Tool | Available Features | Link |
|---|---|---|
| ITK-SNAP [23] | Manual Segmentation | http://www.itksnap.org |
| 3D slicer [17] | Skull stripping, Image registration, Intensity normalization, Segmentation, Visualization | https://www.slicer.org |
| DeepReg [24] | DL based Registration | https://github.com/DeepRegNet/DeepReg |
| ANTs [25] | Image Segmentation, Image Registration | https://github.com/ANTsX/ANTs |
| FSL [26] | MRI Segmentation, Registration | https://fsl.fmrib.ox.ac.uk/ |
| FreeSurfer [27] | Skull Stripping, MRI Registration, Segmentation, Reconstruction | https://surfer.nmr.mgh.harvard.edu |

software tools and in addition there are manual skull stripping software like 3D slicer [17] and algorithms implemented in filter modules in ITK [18].

After skull stripping, the next important steps include the registration, warping and intensity normalization. These steps are essential to make sure the sequences in the exploiting database possess the similar metadata.

The open source software, ITK was used for skull stripping, registration and normalization with histogram matching in [19]. Other than that, the 3D slicer open source software had been used by Wijethilake et al. [20, 21] for skull stripping and registration of the MRI sequences. In addition, a C++ based wrapper function using Insight ToolKit was used by Tixier et al. [22] for the rigid registration of the imaging sequences.

We have presented the tools and the uses of each tool in table 2.1 used in MRI preprocessing.



### Segmentation

In a tumor, there are 3 main pathologically different regions. They are necrosis, enhancement and edema [28]. The next step after preprocessing the images, is

to segment the tumor region, as a whole region, out of the brain tissue. This can also be done as a segmentation into the above mentioned 3 regions too. This segmentation can be performed as a manual task where a specialist or an expert manually annotate the region of interest or as an automated task with computational algorithms. For manual segmentation, there are software such as ITK-SNAP [23] that is widely been used for medical segmentation purposes. However there are several disadvantages associated with manual segmentation. Manual segmentation take time and require a skilled expert. Nonetheless it is a tedious task even for an expert. These segmentations also vary with the annotator. Therefore with the technological advancements more automated approaches were proposed overcoming the above mentioned drawbacks.

The latest automated segmentation approaches focus on deep learning. with the popular convolutional neural networks (CNN) such as fully convolutional networks (FCN) [29] and U-Net [30]. The U-Net [30] architecture has the ability to obtain the spatial and the semantic information with the help of the encoder and the decoder parts. Moreover, the skip connection between the encoder and decoder helps the architecture to learn the fine details. Initially 2D U-Net is used where each slice of the 3D MRI volume is given as the input to the model. However, recently a modified 3D version of this U-Net architecture was proposed, with the ability of handling the whole 3D MRI volume at once [31]. There are recent modification suggested for this U-Net, such as integrating an attention block in the encoder and decoder blocks. Also, further modifications to these attention blocks are proposed by adding skip connections within block for extracting spatial and channel wise dependencies. However, recently the AI researchers claim that the DL algorithms are architecturally saturated and further modification will not increase the models' performance further. They suggest that the research should now be focused on improving the optimization algorithms and modifying the loss functions in order to archive fine segmentation [32].

Brain Tumor Segmentation Challenge (BraTS) is a competition held yearly in parallel with the MICCAI conference. They choose the best performing models for segmentation based on several metrics. In 2018 BraTS challenge, Myro-

nenko [33] has exploited a novel encoder-decoder based 3D architecture, with a small decoder and a comparatively large encoder, that extract imaging features to generate the segmentation. This model comprises of an additional decoder branch for image reconstruction, which is then used to optimize the network, thus providing space to learn with small amount of data. This was placed 1st in the competition in 2018. For the same competition in 2019 Jiang et al. [34] won the first place for the model he proposed with two stage cascaded U-Net, which consists of a two decoders to enhance the model performance. However, still a lot of improvements can be made in the automated segmentation, as it is not quite solid when segmenting tumor regions with close intensities and ambigious boundaries between the healthy normal tissue and the cancerous tissues. Therefore, deep learning techniques used for segmentation are enhanced daily by various research groups, enabling better segmentation performances day by day.

### Feature Extraction

For studying the association with the survival and the images of tumors, the features should be extracted from the segmented regions. These features include the intensity based features, textural features and shape associated features. These can be extracted from each subregion of the tumor and, from the whole tumor region [35]. Intensity based features, includes the mean, variance and skewness of the intensity distribution of the each tumor subregion. For an example intensity in the enhancement region of the tumor, possess a certain mean intensity compared to the other subregions. These might contain valuable connections with the survival of the patient. Shape associated features can be considered as the coordinates of each subregion, length along each axis for each subregion, area and the complexity dimensions such as fractal dimension. Textural features depicts texture wise details such as the entropy, correlation, energy and dissimilarity features of each subregion [36]. These features are mostly extracted from radiographic images at different phases of tumor identification and growth monitoring. Thus, these are known as radiomics. In addition, the recent explorations have

extracted these imaging related features using DL. These are known as deep features and uncover features that are unable to acquire with vision based features extraction techniques.

Summary of several works that have used radiomics, along with the segmentation technique and extracted features, can be found in Table 2.2.

Sanghani et al. [37] have extracted shape features known as bounding ellipsoid volume ratio, i.e. features of ellipsoid that bounds the tumor regions such as orientation, sphericity and spherical disproportion and texture features. Feng et al. [38] have also obtained volume related features for each tumor sub region and fused with resection status, i.e. the surgery to remove the tumor is performed or not, and age for modelling the survival prediction.

Sun et al. [39] have obtained grey level intensity first order statistics features, such as mean, standard deviation , maximum and the minimum intensities and variance of the intensities and the entropy for each sub region of the tumor, i.e. non-enhancing region, edema and the necrosis regions, and then for the whole tumor. In addition, they have obtained axis lengths for the major and minor acis, surface area, surface area to volume ratio, volume features as shape features. As texture features, gray level run length matrix features and gray-level co-occurrence matrix features are also extracted. All these features were extracted using the Pyradiomics library [40] available in Python.

Manually handcrafted imaging features were extracted by Han et al. [41] from the tumor regions and combined with deep features extracted using a VGG-19 [42] DL model, for survival prediction. In addition a CNN [43] was used by Lao et al. [19] to extract features for survival prediction. Nevertheless, they have used handcrafted geometric, intensity and textural radiomic features obtained using a Matlab based feature extraction tool.



### 2.1.3 Genomics

As we mentioned earlier, gliomas arise as a result of genetic changes that occur in the molecular level. These changes can either be genetic or epigenetic, that

Table 2.2: Segmentation methods and imaging features extracted in glioma related studies.

| Related Work | Num. of patients | Segmentation | Features extracted |
|---|---|---|---|
| [39] | 285 | Ensemble model with CNN [44], DFKZ Net [31] & 3D U-Net [45] | First order statistics, shape and texture features |
| [38] | 285 | An ensemble of 3D U-Nets [45] | 6 volume features, age |
| [46] | 285 | VGG-16 [42] based FCN | Age, size and coordinates |
| [37] | 163 | ROI was manually annotated. | Texture, shape and volume features |
| [47] | 163 | ROI was manually annotated. | Histogram features |
| [48] | 163 | Deep U-Net [49] | Texture, volume, euler, histogram |
| [41] | 128 | ROI was manually annotated. | Handcrafted texture features & deep features |
| [50] | 69 | ROI was mannually annotated | clinical information & deep features |
| [51] | 335 | 3D Attention U-Net | geometry, intensity & volume features |
| [33] | 285 | encoder-decoder based 3D architecture | - |
| [32] | 369 | 3D U-Net architecture | - |
| [34] | 335 | Two stage cascaded U-Net | - |
| [19] | 112 | ROI was manually annotated. | Geometry, intensity, texture features & deep features |

can activate the oncogenes, genes that are responsible for cancer development, or inactivate the tumor suppressor genes, genes that suppress the possibility of arising a cancer. With the technological advancements, more specifically in the omics associated high throughput sequencing technologies, the genetic changes in gliomas can be analysed in depth. Consequently, these profiles are also being used in survival related analysis as well, leading for new opportunities and advancements in survival prediction. Therefore, prognosis prediction studies for gliomas have adopted various types of genomic high throughput profiles, such as gene expression, mutation and methylation profiles. These have created a path for more broad analysis. With high utilization and adaptation of genomic profiling in research, the cost for acquiring them have also significantly reduced with the time. This, recently many research institutions have publicly provided these genomic profiles of glioma for research purposes. The Cancer Genome Atlas (TCGA) [1] and Chinese Glioma Genome Atlas (CGGA) [2] are two of such public cohorts.

### Gene expression profiling

Agilent, Illumina and Affymetrix platforms were used initially to obtain the DNA sequences for gene expression profiling using microarray technology. These obtained expression levels are then obtained by comparing the fragments of DNA with a standard target sequence. However, this method is unable to obtain the expression levels of pre unknown genes. Consequently, addressing this drawback next generation sequencing (NGS) was introduced to the sequencing research field. A modern platform used in NGS is Illumina HiSeq. NGS has the ability to provide an extensive profile of the transcriptome by acquiring pre unknown transcripts and the non coding regions. Thus, it is highly appropriate for gene expression profile acquirement [52].

[1] http://cancergenome.nih.gov/
[2] http://www.cgga.org.cn/

### Methylation profiling

DNA Methylation happens in the DNA, by adding a methyl group to the CpG sites as a epigenetic alteration. This function regulates the transcription of genes, differentiating the expression of some genes, by adding methylating the those genes' promoter region. Illumina platforms, are used for acquiring these high throughput methylation profiles and Infinium Human Methylation 450k (IHM-450k) and Infinium Human Methylation 27k are two of the popular BeadChips kits utilized [53].



### Mutation profiling

Mutations that arise in cells within a particular region can be a reason to occur a tumor, which is known as tumorigenesis. If these mutations happen in the tumor suppressor genes, i.e. the genes that are responsible to mitigate tumors by repairing DNA alterations, it can result in uncontrollable cell division. IDH1/IDH2 mutation is one of those mutations that occur in glioma cells, causing for longer survival of glioma patients [54]. There are several types of mutations known as silent, nonsense and missense mutations, which can be recognized through mutation profiling of the DNA/RNA sequences.

Out of these profiling methods, gene expression profiling is widely utilized in risk estimation tool development for gliomas [55, 56]. In addition to that, the associations between these profiles are closely observed to identify the survival related alterations and the genes associated with survival [57].

### Radiogenomics

In addition to explore the associations with the survival and the genomics and imaging features separately, recently a new field has been emerged where these associations are analysed combinedly. This area is known as Radiogenomics and in this, the associations between genomics and images are closely assessed.

In clinical practical, it is much more easy to obtain images of the tumor, as an initial diagnosis tool. Normally after visualizing the tumor region, the

cancerous status of the tumor is identified by performing biopsy. This comes under the second step after initial imaging based diagnosis. Also, it is operative technique and therefore, quite risky. Only at this stage, the clinicians are capable of analysing the genomic profiles of the glioma region. Thus, it is more convenient if the clinicians can get an idea about the genomic status without performing any kind of invasive surgical intervention, by using images. In one of our outcomes of this research [21], we could identify the imaging biomarkers associated with the glioma subtypes, that is categorized based on the genomic level alterations. This will provide a clinicians a better understanding about the genomic level characteristics of the glioma through the images obtained at the initial diagnosis level.

Wijethilake et al. [20] have proposed a radiogenomic based approach for survival prediction of glioma patients, by fusing the imaging features and the gene expression profiles. In addition to that study, Tixier et al. [22] have proposed a method for survival prediction by combining imaging based features such as texture and shape features extracted from the segmented tumor regions, and the gene level alteration information of the O-6-methylguanine-DNA methyltransferase (MGMT) methylation and the IDH1 mutation.

### 2.1.4 Other data types used in Survival Analysis

#### Histopathology

Some researchers also focus on using pathological images, i.e. the images of the tissues that visualize the cells within, for survival analysis. These are also obtained through invasive biopsy or surgery, and the those specimens are closely observed through a microscope to visualize the tissue region. In clinical practise, Histopathologists are trained to analyse those samples to diagnose the glioma type and the status. They observe the effected cell type and decide the glioma type based on the histopathology, categorized as astrocytomas, ependymomas and oligodendrogliomas. Thereby they can roughly predict the prognosis of the tumor.

These pathological images have been used in survival analysis by several researchers. A Deep Learning based method was proposed by Mobadersany et al. [58, 59] for prognosis prediction using the pathological images for glioma patients, while integrating with the genetic biomarkers.

#### Clinical information

As clinical features there are several factors that are been taken into account when performing survival analysis. Age is one of the main clinical factors associated with the survival. Puybareau et al. [46] ranked 1st in the MICCAI BraTS challenge 2018, by predicting the survival only using the age as the input. In addition to the age, there are other clinical factors such as the resection status of the patient used in survival prediction.

### 2.1.5 Public Glioma Cohorts

The above mentioned various data types which can be used in survival studies are publicly available for research purposes. They mainly consist of images, radiological and pathological and genomic profiles, along with the clinical information of each patient case. The National Cancer Institute's (NCI's) Genomic Data Commons (GDC) [3] gives access to the The Cancer Genome Atlas (TCGA) data repository, that includes the genomic, pathological images and the clinical information for various cancer types including glioma, both Glioblastomas and Lower Grade Gliomas. The corresponding radiological images for those patient records can be found at the The Cancer Imaging Archive (TCIA). Additionally, TCIA also enables access to other latest repositories such as the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC), with on going data collection. This dataset also contains the proteomics profiles of different gliomas.

The Multimodal BraTS [4] is another data repository which contains the pre-operative, i.e. before performing surgical resection to remove the tumor, mul-

[3] https://portal.gdc.cancer.gov
[4] https://www.med.upenn.edu/cbica/brats2020/data.html

timodal MRI scans of both higher grade and lower grade gliomas. They also provide the clinical information such as age. This imaging repository is unique, as it provides the manual annotations of tumor sub regions, annotated with the help of neurosurgeons and other experts. Therefore, this is used as a benchmark dataset for segmentation tasks.

There are also cohorts which provides genomic profiles of glioma patients. The Chinese Glioma Genome Atlas (CGGA) is one such cohort with whole exome and mRNA sequencing data, mRNA and microRNA microarray profiles and methylation profiles of high grade and lower grade gliomas. Some researcher have also developed applications and web portals to analyze and download the expression profiles, methylation profiles and DNA mutation landscapes. Another commonly utilized publicly available data cohort is the Gene Expression Omnibus (GEO) [5] and the Ohio Brain Tumor Study (OBTS).

A summary of these datasets and the corresponding studies which had utilized those are briefed in the Table 2.3.

Table 2.3: Frequently used Glioma cohorts for survival analysis of Glioma patients. TCGA: The Cancer Genome Atlas; CPTAC: Clinical Proteomic Tumor Analysis Consortium; BraTS: Brain Tumor Segmentation; CGGA: Chinese Glioma Genome Atlas; GEO: Gene Expression Omnibus, OBTS: the Ohio Brain Tumor Study

| Dataset | Data types | | | | | Studies that used the dataset |
|---|---|---|---|---|---|---|
| | Rad | Path | Geno | Proteo | Clinical | |
| TCGA | × | × | × | | × | [20, 22, 55, 56, 60–62] |
| CPTAC | × | × | × | × | × | [63] |
| BraTS | × | | | | × | [36–39, 46–48, 51] |
| OBTS | | | × | × | × | [62] |
| CGGA | | | × | | × | [55] |
| GEO | | | × | | × | [56, 60, 61] |

[5] https://www.ncbi.nlm.nih.gov/geo/

### 2.1.6 Preprocessing of Glioma survival related Data

#### Data Cleaning

In order to get rid of the noise while enhancing the quality of data, data cleaning is an essential step followed in data analysis tasks. The facts that might lower the quality of data can be recognized are the outliers, the repeated or duplicated entries and the missing information in some entries. More specifically the medical records have this issue with missing data in some records due to the inefficiencies that occur while data collecting and entering for the databases. This is addressed either by considering the mean/mode of the existing other records or by removing such records with missing information. The mean is assigned when the corresponding features is a continuous variable. For an example, if the age of a certain patient is missing in the database, the mean of the rest of the records is considered as the age of that particular record. The mode of the rest of the records is assigned when the correponding feature is categorical. However, literature has proved that the removing records is not an appropriate method to follow as it effects the statistical power and the variance of the corresponding features [64].



#### Normalization

Normalization of features is mandatory to obtain a better fitting data and model. Standardization, also known as Z score transformation, is a popular tool used in forging the data to follow a standard normal distribution with zero mean and unit standard deviation. Min-Max normalization is another common method that normalizes all the features to a similar range, mapping each feature's minimum and maximum values to a given range. Normalization is an important task in any analysis to avoid the influence that some feature makes, compared to the rest [12].

### Dimensionality Reduction

Dimensionality reduction is essential in the medical field to optimize the performance of the learning algorithm, increasing accuracy. According to Ladha et al. [65], Dimensionality reduction is vital to reduce the memory requirements for storage and to increase the algorithm processing speed. Moreover, Redundant, noisy, irrelevant data can be removed with this. Nevertheless, when utilizing the algorithm in real-time or in the testing phase, the most necessary features can be extracted, saving the resources and the time. Dimensionality reduction has two techniques, known as feature selection and feature transformation.

**Feature Selection:**

In feature selection, the features that contribute to the learning process are being selected, removing the redundant features. A subset of original features is chosen, increasing the efficiency and the accuracy of the learning task. This can be done manually or using the algorithms. Feature selection based on algorithms can be discussed under 3 categories, filter methods, wrapper methods, and embedded methods.

As a filter method, correlation-based feature selection is used with high dimensional data as it requires less computational power. Therefore correlation has been utilized in feature selection with high dimensional genomic data [66]. The most commonly used correlation tools are Pearson correlation, Spearman's rank correlation, Kendall rank correlations, and intraclass correlation. Based on the type of data, either continuous or categorical, the correlation tool can be determined. Lao et al. [19] obtained the best feature for the DL based survival prediction model with the intra-class correlation coefficient. Besides, Pearson's correlation coefficient has been often used to identify the relationships between differently methylated genes and differentially expressed genes for prognostic risk score development in glioma patients [67].

Wrapper feature selection methods tend to find the features based on the predetermined learning algorithm, requiring a high computational power compared to the other two. Recursive Feature Elimination (RFE) [68] is a famous wrapper

method used for feature selection in glioma survival prediction. This has been utilized in radiomics based survival prediction along with Support Vector Machine (SVM) [37, 48]. The high time consumption and computational power are significant drawbacks of this wrapper method.

**Feature Transformation:**

In feature transformation, original features are transformed into other sets of significant features. Thus, the information in the original feature space remains in the transformed feature space.

Principle Component Analysis (PCA) provides a clear overview of multivariate data by increasing data interpretability. This method creates new uncorrelated variables, maximizing the variance. PCA is considered as the most widely used dimension reduction method [69].

## 2.2 Glioma Survival Analysis Approaches

Survival analysis of glioma patients is performed using various analysis methods, including ML and statistical analysis based methods. In the past two decades, the attention of the research community has been drawn by the ML techniques widely. As a result, much research was followed to develop a precise survival prediction with learning methods and the features we mentioned in Section 2.1. Other than that, there are statistical approaches, prognostic risk score models, and prognostic nomograms proposed in glioma survival-related studies. In Table 2.4 and Table 2.5, we report feature extraction methods and the analysis methods followed by several studies that have used radiomics and radiogenomics respectively as the input.

### 2.2.1 Machine Learning based Survival Analysis

Different ML algorithms are utilized to predict the survival of glioma patients with a selected/filtered set of features. Throughout the past decade, many ML techniques have been used to predict the overall survival of glioma patients. Not just with radiomics, but also with genomics and radiogenomics, survival predic-

Table 2.4: Feature selection and analyzing techniques used in survival prediction of gliomas with radiomics. IR: Importance Ranking; PCA: Principle Component Analysis; RFE: Recursive Feature Elimination; CC: Correlation Coefficient; RF: Random Forest; LR: Linear Regression; SVM: Support Vector Machine; XGB: eXtreme Gradient Boosting; ST: Survival Tree; ANN: Artificial Neural Network

| Related Work | Feature selection | | | | ML Technique | | | | | | Acc |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | IR | PCA | RFE | CC | RF | LR | SVM | XGB | ST | ANN | |
| [39] | × | | | | × | | | | | | 61% |
| [38] | | | | | | × | | | | | 61% |
| [46] | | × | | | × | | | | | | 61% |
| [37] | | | × | | | | × | | | | 89%(CV) |
| [47] | | | | × | | | | | × | | 67%(CV) |
| [48] | | | × | | | | | × | | | 73%(CV) |
| [50] | | × | | | | | × | | | | 89%(CV) |
| [36] | | | | | | | | | | × | 46.8% |
| [51] | | | × | | | | | × | | | 48.3% |

Table 2.5: Feature selection and analysing techniques used in survival prediction of gliomas with radiogenomics. RFE: Recursive Feature Elimination; LASSO: Least Absolute Shrinkage and Selection Operator; KM: Keplen Meier; LR: Linear Regression; SVM: Support Vector Machine

| Related Work | Feature selection | | Analysing Technique | | | Acc |
|---|---|---|---|---|---|---|
| | RFE | LASSO | KM | LR | SVM | |
| [22] | | × | × | | | – |
| [20] | × | | | × | × | 90% |



tion has been performed utilizing ML tools. Moreover, these techniques can be categorized into supervised, semi-supervised, and unsupervised learning strategies, based on the availability of the survival outcome. Supervised learning models require a clinical outcome or the survival label of each patient case. The supervised model learns to map between the given input and the ground truth label via a pre-defined loss function.

### Support Vector Machine

SVM [70] is a frequently used supervised learning approach for classification and regression tasks. This has been used in overall survival prediction as a classification task, into short, medium, and long survival groups and also as a regression task to predict overall survival in days. SVM is appropriate for high-dimensional

data analysis, due to its ability to overcome the large dimensionality. On the other hand, in SVM, linear models can be extended to non-linear models with kernels by transforming input instance space into a high dimensional feature space. The SVM classification function is as follows.

$$\hat{\Phi}(\boldsymbol{x}') = \text{sign}\left(\langle \mathbf{w}.\boldsymbol{x}' \rangle + b\right) \tag{2.1}$$

where, $\boldsymbol{x}'$, the test point, w, the coordinates of the separating hyperplane and $b$, bias to the origin. With this the SVM classify the samples into classes through hyperplanes in a multidimensional space.

SVM has been frequently facilitated in 2 class (short <400 days and long > 400 days) and 3 class (short <300 days, medium 300-450 days , long > 450 days) survival prediction of Glioma patients with radiomics [37, 50]. Nie et al. [50] obtain deep features adopting a 3D convolutional neural network and combine with clinical features for survival class, long and short survival prediction with SVM. They observe that functional MRI (fMRI) and diffusion tensor imaging (DTI) provide valuable information for survival prediction than T1 MRI. Sanghani et al. [37] use a set of novel shape features that have a high contribution to the survival class prediction after selecting based on RFE. They obtain a high accuracy proving that RFE, along with SVM, is an effective approach for survival prediction.

Emblem et al. [71] obtain histograms of whole tumor relative cerebral blood volume (rCBV) from MRI. The authors have utilized SVM for survival prediction after 6 months, 1, 2, and 3 years after the diagnosis. They have developed 4 separate models that returns the most probable outcomes after 4 time periods. This study has tested the proposed models on independent data and has recognized that their model is insensitive to the treatment changes and image acquisition routine changes. They further demonstrate that the proposed SVM based model can predict the survival with higher accuracy compared to an expert in the clinical field. SVM has also been used for GBM subtype classification [21, 72].

In Support Vector Regression (SVR), the algorithm learns the best fitting

linear function in the feature space, optimizing the epsilon insensitive loss function with a regularizer. However, this is unable to work with the censored data related to survival, as it is unable to identify the event occurrence for censored instances. Thus, Khan et al. [73] has proposed a SVR algorithm for censored data by modifying the epsilon insensitive loss function asymmetrically.



### Survival Trees

As a non-parametric tool, decision trees are widely used in survival analysis due to their flexibility and the ability to handle various data structures. Specifically, it does not require to specify the links between the covariates and the outcome beforehand as it is capable of automatically developing interactions. Basically, in a decision tree, covariate space is split recursively until the nodes of the tree align with the outcome of the covariates. These binary splits are performed with a single covariate. If this covariate $X$ is continuous, split is formed with $X \leq c$, where c is a constant. If this covariate $X$ is categorical, split is based on the criterion $X \epsilon \{c_1, c_2, .., c_k\}$, where $c_1, c_2, .., c_k$ are possible values of $X$. Initially, at the root node, all the covariates are available, and recursive binary splits are performed until the stopping criterion is met. Consequently, this gives an overfitting large subtree. Thus, the most appropriate subtree is chosen with a selection and pruning method. At the terminal node, for classification, the most prevalent class label is chosen, and for regression, the training samples are averaged.

This tree structure is initially applied for survival analysis by Marubini et al. [74] and Ciampi et al. [75]. The difference between the traditional decision tree and the survival tree is the choice of splitting criterion. The aforementioned decision tree does not consider either the interactions between the features or the censoring status. Originally survival trees were proposed with the criterion of minimizing the homogeneity of each node that also minimizes the loss function. Gordon et al. [76] initiated this with the Wasserstein metric between the survival functions, paving the ground for the criterion, based on the dissimilar-

ity between the nodes. Ciampi et al. [77] proposed log-rank statistics, likelihood ratio statistic, and Wilcoxon–Gehan statistic to measure the heterogeneity between the nodes. This method [78] is been utilized for establishing 6 prognostic groups of malignant glioma patients [79]. Moreover, this recursive partitioning analysis approach is being used for overall survival analysis of lower-grade glioma patients based on pre-treatment factors [80]. With all these improvements, recently censored observations are restored with expected survival times, and the median survival tree is developed with L1 loss function [81].

The recent survival studies of glioma patients exploit survival trees with various types of data related to prognosis. Gandia et al. [82] assess the metabolomic profiles acquired from biopsies to predict survival class (short, intermediate, and long OS) of glioma patients with a tree-based model. This classification tree has split up to 3 branches, where the first node criterion is the myo-inosito level and suggests that high myo-inosito levels can cause for longer survival. The second branch cleaves the patients with a low myo-inosito level, based on the glycerol-phosphorylcholine (GPC) level. The final branch splits based on the alanine and glycine levels of glioma patients. However, this study is limited to 46 patient cases; this might have caused overfitting.

Fig 2.2 represents a similar tree developed for survival class prediction based on the gene expression levels. The first branch of the tree is divided based on the expression level of the Solute Carrier Family 30 Member 7 (SLC30A7) gene. The patients with expression level less than -0.067 is split again based on the expression of the YTH domain-containing family protein 2 (YTHDF2) gene.

### Random Survival Forest

Traditional survival trees are considered to be unstable, as it can give different survival functions even for small permutations in the training set. Hence ensemble models, bagging [83] and random forest regression [84] are proposed to overcome the instability by reducing the variance of simple trees.

In the Bagging algorithm, multiple versions of the bootstrap sample [85] are

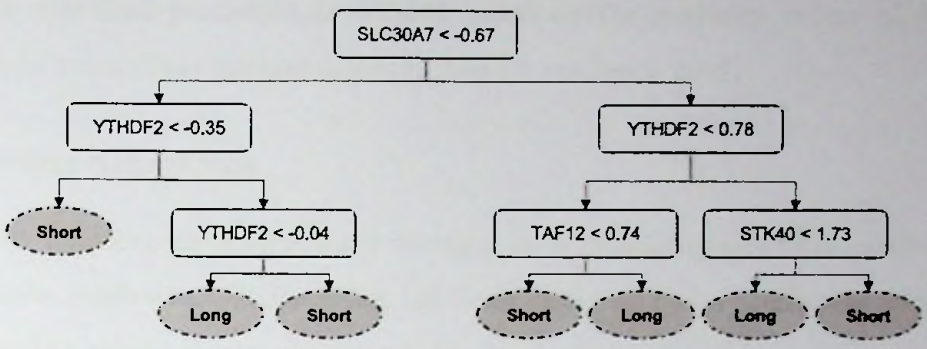



Figure 2.2: Sample survival tree for survival group classification into short, medium, and long survival.

obtained from the given data. Then, for each sample, unpruned trees are built, and the final prediction is acquired by taking the average over all the versions of survival trees. Later this bagging algorithm inspired the subsequent ensemble algorithm, random forests.

Unlike bagging, the splitting criterion of random forests only uses a random subset of all the attributes at the internal nodes, which gives the best prediction. Varieties of splitting criterion have been suggested over the year, such as generating random cutting point for each selected feature and selecting the best among them [86]. This idea is extended by incorporating multiple cutting points for each selected feature and comparing between them [87].

Recently, random forest regression has been widely utilized in radiomic based overall survival prediction [39, 46], as a regressor for overall survival prediction of glioma patients in days. Sun et al. [39] have achieved an accuracy of 61% with the random forest and have considered the limited number of samples might have caused the inadequate efficacy.

Choi et al. [88] have exploited non-imaging features such as IDH status, age, WHO grade, and resection extent for predicting overall survival. Subsequently they incorporate imaging features extracted from MRI for the survival analysis. Thus, they could observe a significant improvement in survival prediction by adding radiomics. Puybareau et al. [46] derive imaging features and build 10 decision trees, each with 3 randomly selected features. As a conventional random

forest, the final prediction is decided based on the majority voting of the 10 decision trees. This method MICCAI BraTS challenge 2018.

### Boosting Algorithms

Lately, boosting algorithms have emerged as a promising ensemble method for accurate prediction. Unlike forest methods boosting has a sequential approach where the model learns to improve the performance of the predecessor. The boosting approach has been applied for neural networks, and later, Freund and Schapire [89] introduced the AdaBoost (Adaptive Boost) algorithm solving previous issues. At first, all the weights are set equally, and each round, the model in series, is weighted, giving priority to the incorrectly classified examples. In gradient boosting machines, in each round, weights are set based on the maximum correlation of the predictors with the negative gradient of the loss function [90]. Later, this gradient boosting algorithm is also modified for censored data in survival analysis [91].

### Linear Regression

Linear regression is a frequently used method for survival data analysis [20]. This is capable of modeling a linear equation with the explanatory variables to output the dependent variable. In survival analysis studies, features that have associations with survival are used as explanatory variables, and the dependent variable is mostly the survival in days from the initial diagnosis or the percentage of survival after a certain time period. However, dealing with censored data without knowing the actual event times is a challenge for linear regression modeling [12].

### Bayesian Methods

Bayesian methods are recognized as a prevailing tool in ML used in both classification and regression tasks. They outperform the other techniques by dealing with uncertainty with the prior knowledge in order to make subtle inference and predictions. Bayesian approaches are also flexible by providing flexible algorithms,

characterising latent structure, uncertainty etc. Bayes theorem is the basic fundamental theorem behind the Bayesian methods. It determines the relationship between the posterior probability and the prior probability, that is affected by particular event that occurred in between. The bayesian posterior probability for a given feature set $\mathcal{F}$, and the model parameter $\Omega$, is given by,

$$p(\Omega|\mathcal{F}) = \frac{p_0(\Omega)p(\mathcal{F}|\Omega)}{p(\mathcal{F})} \tag{2.2}$$

where $p_0(\cdot)$ is the prior probability before the occurrence of dataset $\mathcal{F}$. after training, the model predicts for a given feature set $\mathbf{x}$ [92],

$$\begin{aligned} p(\mathbf{x}|\mathcal{F}) &= \int p(\mathbf{x}, \Omega|\mathcal{F})\mathrm{d}\Omega \\ &= \int p(\mathbf{x}|\Omega, \mathcal{F})p(\Omega|\mathcal{F})\mathrm{d}\Omega \end{aligned} \tag{2.3}$$

There are two Bayesian methods, that are being widely used in the glioma survival prediction [93–95].

**Naïve Bayes (NB) [96]**

Based on the bayesian rule (equation 2.2), for a given feature set ($\mathbf{x}$ the probability of belonging into class $\phi$, the posterior probability of a NB classifier is given by,$P(\phi_i|\mathbf{x})$. Thus, for a binary classifier (with $\phi_1 and \phi_2$ , two posterior probabilies ($P(\phi_1|\mathbf{x})$ and $P(\phi_2|\mathbf{x})$) are calculated and if $P(\phi_1|\mathbf{x}) > P(\phi_2|\mathbf{x})$ then $\mathbf{x}$ belongs to class $\phi_1$ and if $P(\phi_1|\mathbf{x}) < P(\phi_2|\mathbf{x})$ then $\mathbf{x}$ belongs to class $\phi_2$. The class is chosen arbitarily if the posterior probablities are equal. However, in this Naive bayes method assume a mutually independence between the feature set, which is mostly not accurate for survival related data. Also, the complexity is another drawback when it comes to large datasets [97, 98]. Nonetheless, as a white box this is easily understandable for clinicians and other personnel.

**Bayesian networks (BN) [99]**

In Bayesian networks, features are considered as dependent on each other and can be easily interpreted or visualized the relationships between the features. Bayesian Neural Network (BNN) was proposed by Wijethilake et al. [11] for over-

all survival class prediction with mRNA gene expression for glioma patients. The results highlight that the BNN gives a higher accuracy than the other traditional ML techniques such as SVM, RFC and ST.

Zhou et al. [94, 100] leverage the Naive Bayes algorithm to predict the survival classes (long and short) as a classification task. Nevertheless, they also use imaging biomarkers extracted from tumor subregions in GBMs to predict survival time as a regression task. The authors have used distance features that represent the heterogeneity between regions for survival prediction. This study has shown that Naive Bayes outperforms the SVM and K Nearest Neighbor classifiers, with a selected subset of features. However, the deficient dataset of 16 cases is a constraint in this study.

Further Naive Bayes has been utilized by Piccolo et al. [101] for predicting survival status after 2 years from the diagnosis with molecular-level data.



### Artificial Neural Networks

Artificial Neural Networks (ANN) are ML algorithms that resemble the biological neural systems, which were established by Rosenblatt in 1958 [102]. Thus, based on the neuronal activities, the concept of ANN was initially established by McCulloch et al. [103]. A basic ANN model, as shown in Fig 2.3, consists of 3 layers; an input layer that receives the input data, a hidden layer that processes the input data, and an output layer that deliver the results. The hidden layer analyzes the inputs and finds out patterns that give an output based on the associations between the inputs and the outputs. This is done in multiple iterations until the best output is achieved, i.e., the error between the output and the expected outputs has reached a minimum value. Each layer consists of artificial nodes, with weights determined through learning and predetermined activation function. The network retains the learned parameters to predict output for an unseen input. The black box behavior of the ANN is a significant drawback, and also, training an ANN is a time consuming task. However, with the recent developments in the computing technologies, ANN is frequently used in

survival prediction, while extending as deep neural networks. In fact, ANN does not require a linear relationship between input and output to model the behavior.

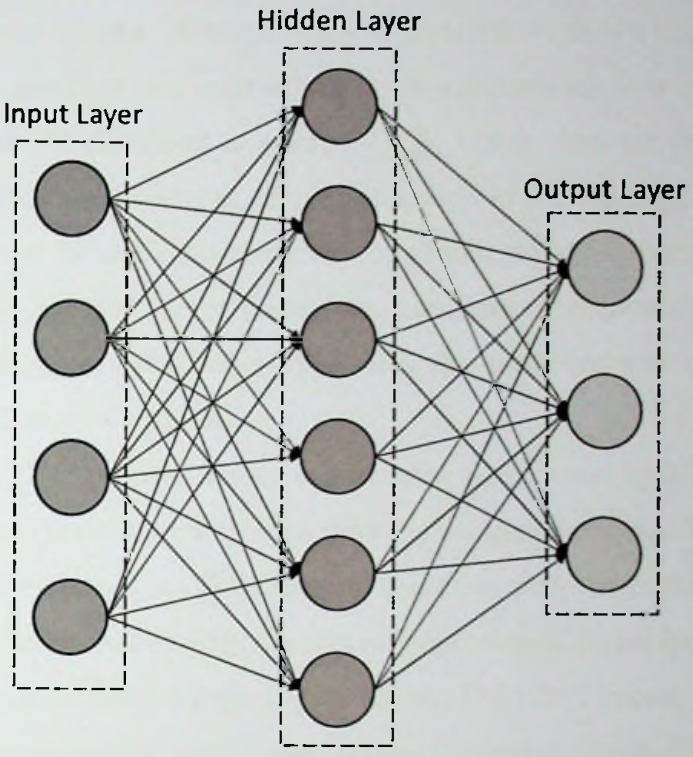



Figure 2.3: Artificial Neural Network consists of a single hidden layer.

In review of the literature, Faraggi and Simon [104] developed an ANN for survival analysis of cancer for the first time in 1995. Recently, neural network is incorporated with the cox proportional model by converting the output layer to a cox regression model, for survival analysis with high throughput omics data in LGG and other types of cancers [105]. Neural Network is utilized in glioma survival prediction to predict the survival time in days directly. Islam et al. [36] propose an artificial neural network to predict overall survival in glioma patients with radiomics. In this work, the input layer consists of the radiomic features such as geometrical features, volume features, texture features extracted from the tumor region, and the output layer is a single node directly predicting the overall survival in days.

### 2.2.2 Deep Learning based Survival Analysis

Recently DL has drawn the attention of the medical researchers with the availability of large amounts of data. Nevertheless, DL models have shown improvements, by boosting the precision and accuracy in many applications, over classical ML paradigms, which we discussed in section 2.2.1. Other than for segmentation and for engineering deep features, direct utilization of DL learning models are not frequently used for outcome prediction in glioma patients. With the recent advancements in digitized image acquisition and data management associated with virtual microscopy technologies, survival is also predicted with pathological images in several studies with DL.

Yousefi et al. [106] propose a deep neural network, followed by a cox survival model for outcome prediction with high dimensional genomic data. In their second work, Mobadersany et al . [58] integrate genomics with the histology images for outcome prediction, where they observe an improvement in performance with both in contrast with histology prediction alone. The CNN known as survival-CNN (SCNN) is inspired by the 19-layer Visual Geometry Group (VGG) architecture that returns a risk, given to a cox proportional hazard layer to optimize the model via back-propagation. Further, the SCNN is trained with histology images to predict the risk, and integrated IDH mutation status and the 1p/19q codeletion, as genomic biomarkers, to train a 3 variable cox regression model.

Rathore et al. [59] also propose an overall survival classification approach with the ResNet architecture. They obtain an accuracy of 84.32% for survival class prediction, long and short survival prediction in glioma patients. Both of these works claim the need for an automated patch extraction methods, to overcome the burdens, such as time that occur with manual patch extraction.

However, a major concern associated with training a DL network is the requirement of large amount of labeled data. This occurs as a result of high number of parameters (millions of parameters) , weights and biases in the convolutional layers, that learn to depict the input features to the patient outcome while minimizing the prediction error. As we mentioned above there are specific set of data

Table 2.6: Related work with DL approaches. TCGA: The Cancer Genome Atlas; CNN: Convolutional Neural Network

| Related Work | Used Data | Datasets | DL algorithm |
|---|---|---|---|
| [59] | Pathology images | TCGA | Resnet |
| [58] | Pathology images & Genomics biomarkers | TCGA | Survival CNN |
| [106] | Genomics profiles | TCGA | Deep Survival Neural Network |

augmentation techniques followed by DL researchers in medical field to overcome the data deficit. Randomized rotations and flipping of the training data, contrast and brightness transformations are a few of those methods. Thereby the training cohort can be expanded to train the DL network to obtain a better performance.

Moreover, in spite of the intra-tumor heterogeneity, sampling regions of interests (i.e. patches) from the pathological images allow the DL models to discern the variations effectively. This heterogeneity is clearly visible in pathological images, and thus, Mobadersany et al. [58] have selected the regions of interest from each slide with the help of an expert in the field.

### 2.2.3 Statistical Analysis tools for Survival Analysis

There are three types of statistical analysis methods for survival analysis. *i*) Non-parametric *ii*) Semi-parametric and *iii*) Parametric [12]. Kaplan Meier and Nelson Aalen are the most popular non-parametric methods used in survival analysis, despite the difficulties in interpretation. All these tools are frequently utilized in glioma survival studies, to identify prognostic genes, to visualize the survival function variations between different categories or for comparisons.

#### Survival Data and Censoring

The survival analysis explores the effect of individual covariates and the time until an event, such as death or a specific state, reaches. This includes the partially observed data, known as censored data, which makes survival analysis different from ML. Correspondingly, if an event, such as death, occurs within

the monitoring time period, it is considered as an uncensored record. Censored records do not experience an event within the time period and do not know the exact event occurrence status that might or might not have happened after the time period [12].

### Kaplan Meier Analysis

In survival analysis, there are two functions that depend on the time.

- Survival function: Probability of surviving at least up to time $t$; (pr($T \geqslant t$)).
- Hazard function: Conditional probability of dying at time $t$, if the patient survived until time $t$.

Keplen-Meier Curve [107] is used to evaluate the survival function based on the observed survival times, without considering the underlying probability distribution. The probability of surviving $t'$ time is the cumulative probability of surviving each $t'$ time periods from the beginning of the study.

$$S(t') = p_1 \times p_2 \times p_3 \times \ldots \times p_{t'} \tag{2.4}$$

where, $p_1$ is the probability of surviving the first time period etc. The probability of surviving each $i$ time period and up to $i$ time period is obtained by,

$$p_i = \frac{k_i - d_i}{k_i} \tag{2.5}$$

$k_i$ is the number of alive cases at the beginning of the $i$ time period, and $d_i$ is the number of deaths within the $i$ time period. In the Kaplan-Meier analysis, the data where the censoring occurred just before the $i$ time period are excluded from the $r_i$. Also, the time that the censoring occurred has a probability of survival of 1.

### Log rank test

Log rank test is a statistical hypothesis test initiated to compare two survival curves. The test null hypothesis is that there is no difference between two survival curves (two population groups).



$$\chi^2(\log\text{rank}) = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \tag{2.6}$$

where $E_1$ and $E_2$ are the expected events in 1 and 2 groups and $O_1$ and $O_2$ are the observed events in 1 and 2 groups. Despite the fact that the log-rank test can be used to explain the difference between survival curves of different groups, it does not account for any other variables that can affect the survival curve [12].

### Cox Proportional Hazard Analysis

Cox proportional hazard model [108] is semi parametric survival analysis approach and a multiple regression model, that can account many factors at once for the analysis. Hazard function $\lambda(t)$, the probability of dying at a particular time if survived upto that time, is the dependent variable of this model.

$$\lambda(t) = \lim_{\Delta t \to 0+} \frac{\text{pr}(t \leqslant T < t + \Delta t | t \leqslant T)}{\Delta t} \tag{2.7}$$

Now, for a given instance $i$, the available covariate vector is $\mathbf{z}_i = (z_{1i}, \ldots, z_{pi})$. Thus, the cox model follows the following hazard function.

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\mathbf{z}\beta) \tag{2.8}$$

where the baseline hazard function $\lambda_0(t)$, is a non negative function of time for $\mathbf{z} = 0$. $\beta$ is the coefficient vector associated with each covariate, $\beta^T = (\beta_1, \beta_2, \ldots, \beta_p)$. Between any two instance the hazard ratio is,

$$\frac{\lambda(t, X_1)}{\lambda(t, X_2)} = \frac{\lambda_0(t) \exp(X_1\beta)}{\lambda_0(t) \exp(X_2\beta)} = \exp[(X_1 - X_2)\beta] \tag{2.9}$$

This demonstrate that hazard ratio, is a constant and does not depend on the baseline hazard function $\lambda_0(t)$. Further, the baseline hazard function is the same for all the subjects. Thus, the cox model is a proportional hazard model. The hazard ratio ($HR$), $\exp \beta_i$ is obtained for each feature, and based on that hazard ratio, the effect of those features for the survival is assessed. The $HR > 1$, i.e., the value of $\beta_i > 0$, implies that the $i^th$ feature is positively associated with the hazard function and thus, decreases the overall survival. Cox proportional hazard model is used in glioma survival estimation in many ways. Recently, the cut off value of $HR$ is for feature selection.

Zuo et al. [55] select genes for risk signature development, based on the $HR > 1$, obtained from the univariate Cox model and further, to obtain the genes with the highest impact, the multivariate cox model is utilized. Hsu et al. [109] also utilize univariate cox model for identifying survival-related genes.

### 2.2.4 Other methods used in survival prediction

#### Risk Score Model

Risk score modelling is another common approach for risk estimation of glioma patients. Thus by this, the patients are separated in the high risk and low-risk groups based on the expression or methylation of genes or other features associated with the survival. To obtain the most prominent genes associated with survival, univariate and multivariate cox proportional hazard analysis are frequently utilized [55–57]. The features are filtered based on the Hazard ratio (typically HR > 1), and the p-value (for a significant coefficient p values should be < 0.05). Thus, based on the chosen gene signature, the risk score is calculated, basically based on the following formula.

$$\text{Risk score} = \sum \text{status}_{gene} * \text{coef}_{gene} \tag{2.10}$$

where $\text{coeff}_{gene}$ is the coefficient obtained from the Cox PH hazard analysis for the corresponding gene and $\text{status}_{gene}$ is either the expression level or methylation status of the same gene. The median risk score of the dataset is considered as the

threshold of high and low-risk separation of patients. In order to give a better understanding of this clustering, heatmaps, Kaplan Meier plots are visualized.

Zuo et al. [55] have used two publicly available GBM cohorts, CGGA and TCGA, to identify the prognostic genes. They have initially performed univariate cox regression analysis on both datasets separately and, based on cut-off values of $P<0.01$ and $HR>1$, filtered 49 prognostic genes common both cohorts. Thereafter, the step-wise multivariate cox regression analysis is performed to obtain the 6 genes, D79B, MAP2K3, IMPDH1, SLC16A3, MPZL3, and APOBR as a 6 gene risk signature. This work has identified this 6 gene risk score as an adverse prognostic risk factor compared to the other clinical factors such as chemotherapy, radiation therapy, and age. Since they have used two cohorts, they claim that they have been able to develop a more stable and reliable model relative to the risk score model developed with a single cohort.

A study carried out by Xian et al. [56] has extracted non-coding gene expression profiles for glioma patients for the risk score development, since previous studies have shown non-coding RNAs cause cancer progression.

Apart from gene expression profiles, methylation profiles are also used for developing risk score models. Other than that, radiomics are also utilized for risk estimations. Beig et al. [63] extract radiomic features from T1 weighted MRI to develop risk score models, separately based on the gender for accurate risk estimation. In Table 2.7, we present the summarize of several related works that have developed prognostic risk score models.

### Nomograms

Similar to the risk score prognosis model, another survival estimation tool is the prognostic nomograms. These are generated to estimate the probability of survival after a specific time period of the diagnosis, based on the factors such as gene expression of a selected set of genes. Wang et al. [110] develop a 5 gene prognostic nomogram for GBM patients, considering the gene expression levels of OSMR, BICDL1, SH3BP2, MSTN, and RGS14 genes. Similarly, Gittleman et

Table 2.7: Related work on prognostic risk score calculation. TCGA: The Cancer Genome Atlas; CGGA: Chinese Glioma Genome Atlas; GEO: Gene Expression Omnibus; CPTAC: Clinical Proteomic Tumor Analysis Consortium; TCIA: The Cancer Imaging Archive; LGG: Lower Grade Glioma, GBM: Glioblastoma, Cox PH: Cox Proportional Hazard model

| Related Work | Glioma type | Datasets | Feature selection method | Features used |
|---|---|---|---|---|
| [55] | GBM | TCGA & CGGA | Cox PH | Gene Expression data |
| [56] | Glioma | GEO & TCGA | Cox PH | Gene Expression data |
| [57] | LGG | - | Cox PH | Gene Expression data & Methylation data |
| [60] | GBM | TCGA & GEO | log rank test | Methylation data |
| [61] | GBM | TCGA & GEO | Cox PH | Gene Expression data |
| [63] | GBM | CPTAC & TCIA | Cox PH | Radiomic features |

al. propose a nomogram for LGG patients, considering the Grade of the glioma (either grade II or grade III), Sex, Karnofsky performance status, molecular subtype. Neutrophil/lymphocyte ratio, age, the extent of the resection, and the histology (GBM or LGG) are taken into account by Yang et al. [111] to develop a comprehensive nomogram for glioma patients.

An example of a prognostic nomogram is shown in Figure 2.4. In this, the expression levels of SLC30A7, YTHDF2, TAF12, and STK40 genes are considered to determine the survival probability after 12 months and 18 months. The points given to each gene is determined by the expression level that lies between 0 and 1, and then the total of those points are used to determine the survival probability. However, as we can see, higher the total points cause shorter survival according to the given nomogram.

A summary of related works that have proposed prognostic nomograms for survival probability estimation is given in Table 2.8.

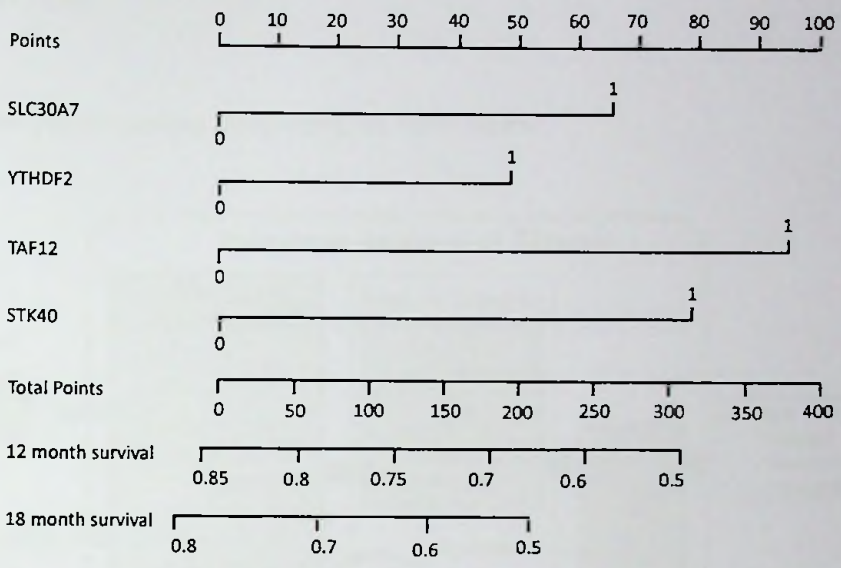



Figure 2.4: A sample nomogram.

Table 2.8: Related work on prognostic nomograms development. TCGA: The Cancer Genome Atlas; CPTAC: Clinical Proteomic Tumor Analysis Consortium; GEO: Gene Expression Omnibus, OBTS: the Ohio Brain Tumor Study, LGG: Lower Grade Glioma; GBM: Glioblastoma

| Related Work | Glioma type | Datasets | Features used |
|---|---|---|---|
| [110] | GBM | TCGA & GEO | Gene Expression data |
| [62] | LGG | OBTS & TCGA | Clinical & molecular factors |
| [111] | Glioma | private dataset | Hematological & clinicopathological factors |

# Chapter 3

# METHODS

## 3.1 The Approaches included in this work

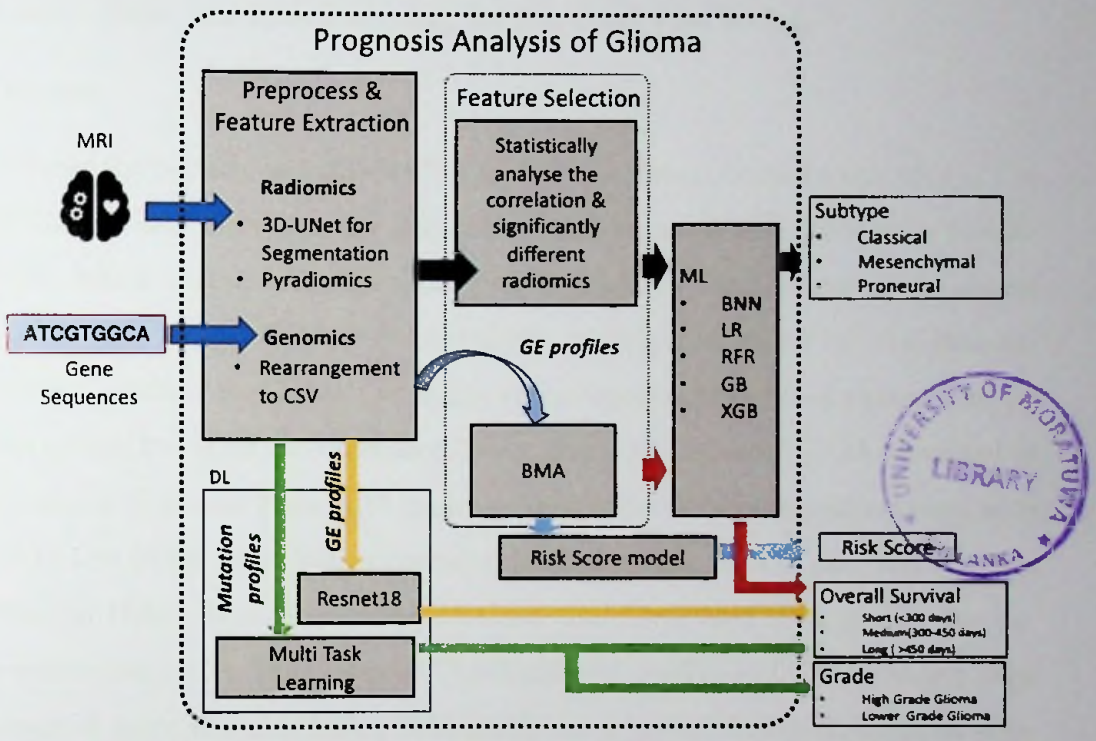



Figure 3.1: Graphical Abstract.

For prognosis analysis we have followed different approaches which we extensively discuss in Chapter 3. The different approaches we have followed are clearly visualized in the Figure 3.1. The different colored arrows depict the approaches we have followed in the following sections.

The Section 2.2.1 consists of two main sub sections, 3.2.1 and 3.2.2, each consists of Machine Learning and Deep Learning approaches followed for survival analysis with Gene Expression profiles in this work. The Sub section 3.2.1 consists of two parts, shown in the brown and light blue arrows, where the each respectively represent the sub section 3.2.1 and the sub section 3.2.1. Next, the

Sub Section 3.2.2 consists of the DL approaches followed with Gene Expression profiles in this research. This is shown in the yellow color arrows.

The path followed in Section 3.3 and Section 3.4 are represented by the green color arrows and dark blue color arrows respectively.

## 3.2 Prognosis prediction with Gene Expression profiles

### 3.2.1 Basic Machine Learning for Survival Prediction

#### Dataset

We used the publicly available TCGA and CGGA data cohorts for this study. The TCGA consists of two glioma sub cohorts known as TCGA-GBM and TCGA-LGG, which contains Glioblastoma (WHO grade IV) and Lower grade glioma (WHO grade II and III) cases respectively. In this section, we use the gene expression profiles, which has the ability to catergorize the glioma patients in the risk groups based on the expression levels. Each TCGA and CGGA consisted of 252 and 315 patient cases with clinical information such as overall survival with 16510 and 24326 genes with expression levels for each patient case respectively. Illumina HiSeq RNA Sequencing platform had been used in acquiring the expression profiles for TCGA cases. The expression profiles of CGGA dataset were obtained using Illumina HiSeq 2000 platform, a powerful high-throughput gene sequencing system. The downloaded expression values were in the gene length into fragments per kilo-base per million mapped reads (FPKM) [112] valued and downloaded in a comma-separated values (CSV) format.

$$\text{FPKM} = \frac{\text{Total fragments mapped reads million}}{\text{Exon length kilobase pair}} \tag{3.1}$$

Thereafter, the downloaded FPKM valued datasets were then log transformed and median centered normalized. There were 13094 common genes in both TCGA and CGGA datasets, these genes were filtered for the analysis. Then, CGGA data were splitted into train and validation sets, while keeping the TCGA data for the testing. Since this is a classification task, the continous variable, Overall survival

in days, were divided into three classes, known as Short survival: overall survival in days $< 300$, Medium survival: $300 <$ overall survival in days $< 450$, Long survival: overall survival in days $> 450$. These are the three classes usually used in glioma survival prediction tasks. The class distribution for each long, medium and short classes are shown in the Table 3.1 in both TCGA and CGGA folds. As we mentioned, the CGGA was splitted into 4 folds, with similar class distribution ratio between train and validation in each fold, while without overlapping the cases.

Table 3.1: Dataset Description

| Class | CGGA dataset | | TCGA |
|---|---|---|---|
| | *Training* | *Validating* | *Testing* |
| Short | 63 | 21 | 68 |
| Medium | 29 | 9 | 44 |
| Long | 145 | 48 | 140 |

### Prognostic Gene Identification

To develop a robust learning model, we had to choose the most prominent genes that reflect the survival of patients using a feature selection method. For this we used the Bayesian Model Averaging (BMA) [113], which has the ability to overcome the model uncertainty via considering the mean of the posterior probability distributions of pre-assumed models. Thus, the posterior probability of being $\Phi$ for the given the training dataset $D$ was obtained using the following equation.

$$\Pr(\Phi|D) = \sum_{i \in S} \Pr(\Phi|D, M_i) \cdot \Pr(M_i|D) \tag{3.2}$$

where, $M_i$ is a given model in the subset, $S = \{M_1, M_2, \ldots, M_n\}$. Initially the genes were ranked based on Cox proportional hazard analysis (Cox-PH), which has the ability to deal with censored data. Cox hazard function (3.3) for a given covariate vector of subject $p$, $\mathbf{z}_p = (z_{1p}, \ldots, z_{ip})$, depicts the probability of dying at a given time $T$, if the patient survived until time $T$.

$$\lambda\left(\mathrm{T}, \mathbf{z_p}\right) = \lambda_0(\mathrm{T}) \exp\left(\mathbf{z_p}\alpha\right) \tag{3.3}$$

where baseline hazard function is $\lambda_0$ and the coefficients of the each $(z_{1p}, \ldots, z_{ip})$ covariates is given by $\alpha = (\alpha_{1p}, \ldots, \alpha_{ip})$. Since the baseline hazard function is same for a single case, for different time $T$, it was neglected. Thus, an approximation for $\alpha$ was required and it was calculated with the following partial likelihood association.

$$\mathrm{PL}(\alpha) = \prod_{p=1}^{n} \left( \frac{\exp\left(\mathbf{z_p}\alpha\right)}{\sum_{\ell \in R_p} \exp\left(\mathbf{z}_\ell \alpha\right)} \right)^{\delta_x} \tag{3.4}$$

In this Equation (3.4), $R_p$ is the risk set, subjects that have not experienced an event by the time $t_p$ and $\delta_i$ is the event status (censored or not) of patient $x$. The $\theta$ parameter was obtained by maximizing the partial likelihood, and thus, according descending order of log likelihood of those values, genes were ranked.

The top ranked 25 genes were assigned to a window and traditional BMA algorithm was applied for survival analysis. Based on the posterior probabilities of those genes, the genes with low posterior probabilities ($< 1\%$) were eliminated retaining the genes with high posterior probabilities. The window was moved along the top ranked genes until 7 genes with high posterior probabilities were obtained. These 7 genes and their posterior probabilities are given in the Table 3.2. Accordingly, TXLNA gene had the highest posterior probability and the other genes had posterior probabilities between 1 and 100%.

Table 3.2: Selected 7 genes and their corresponding posterior probabilities

| Gene | Posterior probability |
|---|---|
| TXLNA | 100.0 |
| WDR77 | 50.5 |
| TAF12 | 48.9 |
| STK40 | 35.9 |
| YTHDF2 | 14.7. |
| SNRNP40 | 8.0. |
| SLC30A7 | 2.0 |

### Survival Prediction

In this section, we follow two approaches for prognosis prediction. First, we used the traditional machine learning algorithms as a classification task to predict the survival class. The chosen 7 genes, using the BMA algorithm based gene identification were used as the input for all the models. One of the issues, we had to tackle during the model training is the class imbalance. For this we randomly sampled the minority classes in order to match with the majority class. In our case we randomly sampled the short and medium class cases in the training split.

#### Decision Tree

Decision tree (DT) [74] is a tree based structure, where each node in the tree specifies a condition of a input covariate that decides the distribution or the final class. The decision tree is fine tuned to obtain the best results, with a maximum depth of the tree to be 4 and the minimum samples required to make a decision at a node to be 10.

#### Random Forest Classification

Random Forest Classifier (RFC) [84] is a widely used ensemble model, that outperforms a single decision tree by overcoming the instability and the variance, with multiple decision trees. Hence, in RFC the number of trees in the forest are tuned to 100.

#### XGBoost

XGBoost [114] is an ensemble model but each tree is trained in a sequential manner, boosting the prediction accuracy. The fine tuned XGBoost model, consists of 50 trees trained sequentially, with a maximum depth of a tree, 4 layers.

#### CatBoost

Catboost [115] is a sequentially trained boosting algorithm, which outperforms the existing boosting algorithms with its efficiency. We utilize multi-class loss function with a learning rate of 0.005 for 2000 epochs for 4 layered symmetric tree.

#### Support Vector Machine

Support Vector Machine (SVM) [70] is a common machine learning algorithm

that is been used for survival prediction of glioma patients. Thus, we explore a support vector classification based on SVM with a linear kernel.

Moreover, we proposed a novel Bayesian neural network based on deep probabilistic programming. Deep probabilistic programming uses the power of deep neural networks with probabilistic models, to enhance the performance while optimizing the related costs [116]. The traditional more frequently utilized algorithms show less performance, as discussed in the results section. Therefore, a need for a better performing algorithm, which can recognize the uncertainties in the predictions, occurred.



**Proposed Bayesian Neural Network**

We proposed a probabilistic programming based Bayesian neural network (BNN) for the survival class prediction of glioma patients. The first layer receives the chosen 7 features as the input. The final layers contains the predicted class, short, medium or long. In the middle the architecture consists of 2 layers comprised with 24 and 12 neurons in each layer respectively. Both hidden layers are comprised of 50% dropout with $1^{st}$ hidden layer followed by a Tanh activation and the $2^{nd}$ hidden layer followed by a rectified linear activation unit (ReLU). All the nodes are fully connected with nodes in the adjacent layers. BNN differs from the traditional artificial neural network by assigning distributions for all the parameters, weights and biases instead of a single value. Mean and scale parameters of all the distributions are initialized to 0.01 and 0.1 respectively. Stochastic gradient descent optimization with learning rate 0.001 is utilized for training the BNN. Further, stochastic variational inference optimizes the trace implementation of evidence lower bound (ELBO) in order to diverge probability distribution of parameters. The proposed BNN architecture is shown in Fig. 3.2.

Implementations were developed with Pyro (version 1.3.1) probabilistic programming language with pytorch (version 1.5.0) [117]. For the other machine learning algorithms Scikit-learn library [118] was utilized. To measure and compare the performance of the applied machine learning algorithms, 4 metrics Accuracy, Precision, Recall and F1 scores were occupied.

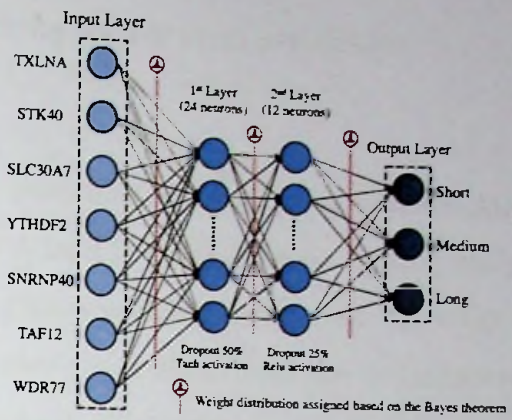



Figure 3.2: Proposed Bayesian Neural network architecture.

### Prognostic Risk model Construction

Further, in order to estimate a risk, we calculated a risk score with the survival related genes of glioma patients. We used univariate cox proportional hazards regression analysis to evaluate the association between the 7 genes obtained from the iterative Bayesian Model Averaging algorithm and the survival status and time. Typically the genes with p values $< 0.01$ and the hazard ratio (HR) $> 1$ are considered as the criteria for them to be candidate genes for the survival estimation. The obtained $\beta$ values, i.e. the regression coefficient obtained from the cox analysis were used to calculate the risk score along with the expression values of the corresponding genes ($\exp_i$). The univariation cox analysis was performed with the R package, survival (version 3.2-3) [119]. The Equation (3.5) indicates the prognostic risk score formula used to obtain the risk score prognostic model.

$$\text{Risk score} = \sum_{i=1}^{n} \beta_i * \exp_i \tag{3.5}$$

$n$ is the number of genes chosen to be included in the prognostic signature. After obtaining the risk score for the CGGA cohort, the median value of the prognostic risk score was used to separate the patients in the high and low risk groups. To clarify, the performance of the proposed signature was validated on the TCGA cohort.

### 3.2.2 Deep Learning for survival prediction

#### Dataset

We utilized publicly available Chinese Glioma Genome Atlas (CGGA) [120] and the Cancer Genome Atlas (TCGA[1]) datasets for our study. The CGGA cohort is composed of 325 glioma patients, along with GE levels of 24326 genes. Out of 325, 315 patient cases are available with the overall survival information. The TCGA dataset comprises 685 patient cases; however, overall survival data is available only for 252 patients. The TCGA dataset contains GE levels of 16509 genes. The survival knowledge for both databases is given by survival in days and further grouped into short-term (<300), medium-term (300-450), and long-term (>450), as shown in Figure 3.3. As shown in the figure, both CGGA and TCGA present class imbalance problem.

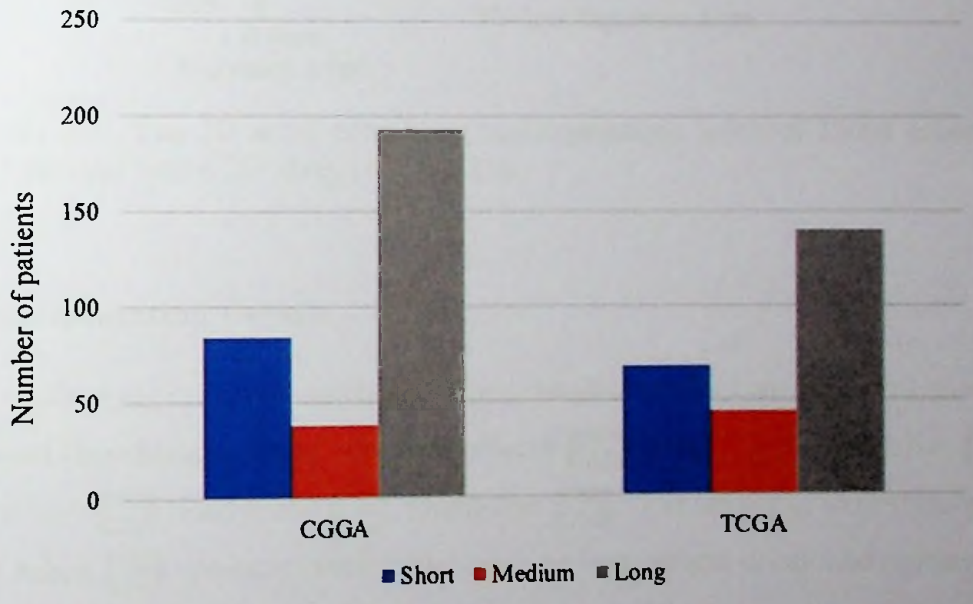



Figure 3.3: Survival Time Distribution (in days) into three class samples - short-term, medium-term, and long term for CGGA and TCGA dataset.

The GE profiles exhibit Fragments Per Kilo-base Per Million (FPKM) values [112]; thus, both datasets were $log_2$ normalized and median centered before analysis. Consequently, 13094 common genes in the TCGA and CGGA datasets

[1]https://www.cancer.gov/tcga

with expression profiles were separated for the deep learning-based survival prediction task. Furthermore, we transformed patient samples from a 1D array of size 13094 into a 2D array of 116 x 116 with zero padding, as shown in Figure 3.4. Finally, we normalized each sample to [0, 1] without any information loss, and feed to the CNN models.

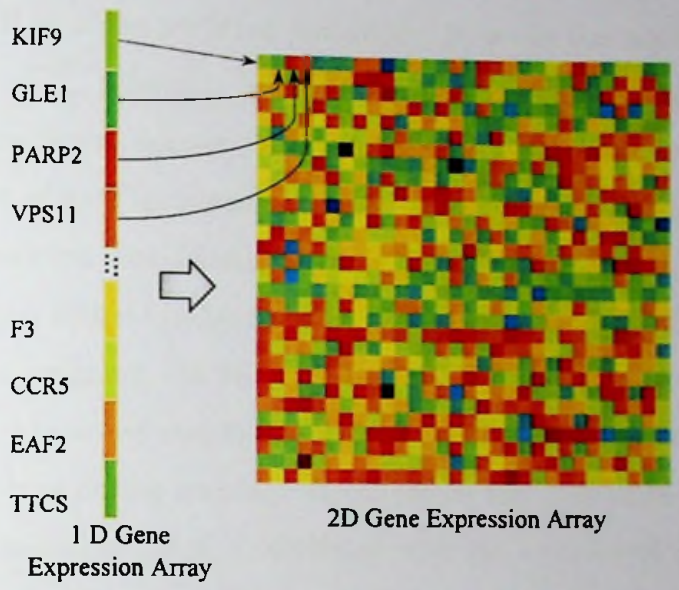





Figure 3.4: The 1D array containing the expressions levels of 13094 genes are transformed into a 2D array of 116 x 116.

### Implementation Details

To conduct experiments, we chose Pytorch vanilla architecture of several state-of-the-art classification model such as ResNet18 [121], VGG16 [42], GoogleNet [122], MixConv [123], ResNest50 [124], ShuffleNet [125]. The networks were trained using Adam [126] optimizer with initial learning rate, weight decay and momentum of $10^{-4}$, $10^{-6}$, 0.9 respectively. We performed 4-fold cross-validation on the CGGA and TCGA with same experimental hyperparameters. A single NVIDIA GTX 1080 Ti GPU was used to conduct all the experiments.

### Methods

In this section, we discussed in detail the formulation of our novel Label Smoothing Focal (LSF) loss and the subsequent variation. We initiated the proposition

by relating cross-entropy loss, focal loss, and their modified version with label smoothing in terms of model calibration and uncertainty.

### Background

The most common loss function for classification task is cross-entropy (CE) as it corresponds to maximizing the likelihood or the probability of a multinomial distribution. If $\hat{p_c}$ is the predicted probability, $p_c$ is the one hot target then CE loss, $\mathcal{L}_{\text{CE}} = -\sum_{c=1}^{N} p_c log(\hat{p_c})$. However, CE loss is unable to tackle with class imbalance issue in the dataset. Focal loss [127] addresses the issue by focusing on a sparse set of hard instances and suppressing the much prevalent easy negatives. It incorporates a modulating factor $(1\text{-}\hat{p_c})^{\gamma}$ to curtail contribution from easy examples, paired with a tunable focusing parameter $\gamma$ to adjust the rate of corresponding down weighing, can be formulated as. $\mathcal{L}_{\text{FL}} = -\sum_{c=1}^{N}(1-\hat{p_c})^{\gamma} p_c log(\hat{p_c})$. Moreover, an $\alpha$-balanced adaptation of the focal loss takes into account the overall imbalance bias during training. It elucidates the imbalance by introducing an arbitrary weighting factor $\alpha$ correlated with the fraction of imbalance for a particular class.

$$\mathcal{L}_{\text{fFL}} = -\sum_{c=1}^{N} \alpha(1-\hat{p_c})^{\gamma} p_c log(\hat{p_c}) \tag{3.6}$$

On the other hand, Cosine loss [128] integrates the previous perception under hierarchy-based embeddings, hence elevating the accuracy for the classification task.

$$\mathcal{L}_{\cos} = \sum_{c=1}^{N} 1 - \langle \varphi(p_c), \psi(\hat{p_c}) \rangle \tag{3.7}$$

Ultimately, equation (3.7), draws the cosine similarity between the ground-truth class embedding, one-hot vector $\varphi(p_c) = p_c$ and its corresponding $L^2$-normalized prediction $\psi(\hat{p_c})$, where $\psi(z) = \frac{z}{\|z\|_2}$.

Despite enhancing cost function to increase the accuracy, the model is poorly calibrated. There are evidences that label smoothing (LS) prevents the over-confidence in model learning and improve the probability calibration [129, 130]. LS

produces the soft target by re-weighting hard targets with a uniform distribution during CE loss calculation. It can be presented as,

$$\mathcal{L}_{\text{LSCE}} = -\sum_{c=1}^{N} p_c^{LS} log(\hat{p_c}) \tag{3.8}$$

where, soft target, $p_c^{LS} = (1-\epsilon)p_c + \frac{\epsilon}{N}$ generates by integrating the labels with uniform distribution of $\frac{\epsilon}{N}$ and $\epsilon$ is the smoothing factor which value consider to 0.1 for the best performance.

**Label Smoothing Focal Loss (LSF)**

We proposed Label Smoothing Focal (LSF) loss by integrating label smoothing (LS) with $\alpha$-variant focal loss (FL). LS prevents over-confidence model prediction by flattening the true probability during training, and FL addresses the class imbalance problem by down-weighting the easy samples and focusing on the hard sample. Therefore, LSF enables better calibration and uncertainty in model learning and boosts performance by addressing the unbalance class issue. We can define the LSF cost function as,

$$\mathcal{L}_{\text{LSF}} = -\sum_{c=1}^{N} \alpha(1-\hat{p}_c)^{\gamma} p_c^{LS} log(\hat{p_c}) \tag{3.9}$$



In equation (3.9) $\alpha$, $\gamma$, $\epsilon$ and $\hat{p}_c$ is the weighting value for class imbalance bias, modulating factor, label smoothing factor and the predicted class probability, respectively. Henceforth, as ($\alpha$, $\gamma$, $\epsilon$) $\rightarrow$ (1, 0, 0), the proposed LSF maps equivalence to cross-entropy loss. After the methodical calibration of LSF parameters, the function yields better model assessment and perception, coupled with the increased capability to penalize easy negatives.

**Normalized LSF Loss (normLSF)**

Cosine Loss [128] uses $L_2$-normalized prediction and target as a regularizer to improve the model performance, specifically datasets with limited sample multi-class distribution. Consequently, We integrated $L_2$ normalization in LSF to form normalized Label Smoothing Focal (normLSF) which normalize the model output and smoothened targets. If $\psi$ is the $L_2$-normalization the *normLSF* can be

formulated as,

$$\mathcal{L}_{\text{normLSF}} = -\sum_{c=1}^{N}(1-\hat{p_c})^{\gamma}\psi(p_c^{LS})log(\hat{p_c}) \tag{3.10}$$

## 3.3 Prognosis prediction with Mutation profiles

### 3.3.1 Dataset

The Cancer Genome Atlas (TCGA) and the Chinese Glioma Genome Atlas (CGGA) datasets were used for this analysis. The whole exome sequencing profiles were downloaded from the TCGA for TCGA-GBM and TCGA-LGG cohorts and the somatic non-silent, i.e. missense and nonsense mutations, were selected and rearranged to a CSV. Similarly, CGGA dataset provided whole exome sequencing, somatic mutation profiles. After filtering the common genes in both CGGA and TCGA cohorts, 7484 genes were used in this study. These profiles are available for 294 and 273 patient cases in TCGA and CGGA respectively. Each patient record consist of the overall survival, which was categorized into 3 classes, Long survival: Overall survival $>$ 450 days, Medium Survival: 450 days $>$ Overall survival $>$ 300; Short survival: Overall survival $<$ 300 days; and WHO glioma grade, which was categorized into 2 classes, High Grade Glioma (Glioblastoma): WHO grade 4 gliomas; Lower Grade Glioma: WHO grade 2,3 gliomas. The TCGA-GBM is considered as the higher grade and the TCGA-LGG is considered as the lower grade gliomas. These two were the two ground truths, we used in training the multi task model. The WHO grade 1 glioma cases were removed as they were mostly non malignant, has a longer survival rate. Thus, the two tasks are grade prediction and survival class prediction. The cohorts were divided into 4 folds, while maintaining the class balance among folds for both grade and survival classes.

### 3.3.2 Methods

Multi task learning has drawn the attention of the Deep Learning researchers recently, as it enhances generalized performance of multiple related tasks. This learning paradigm also addresses the data sparsity problem, by aggregating similar tasks in spirit of data augmentation [131]. Therefore, in our multiple classification task the survival class and the glioma grades are related, as the WHO grades are mostly given based on the survival of the glioma patient. In glioma management, the WHO grade is assigned at the initial diagnosis stage, where the histopathology is closely observed, and then, based on that and the other clinical factors such as age, survival is estimated by the physicians. A model known as MLT-SA has been proposed for classifying the censor status and the survival prediction tasks for survival analysis [132]. However, Multi task learning has lot of room left for improvement in this domain of prognosis analysis of gliomas, which can ultimately provide a clinically applicable tool to support the clinicians decision making process.

In this section, we proposed a Resnet18 [133] based MTL model, for class based overall survival prediction and glioma grade prediction. We introduced GradNorm Asynchronous Task Aware Optimization method, a novel optimization approach for MTL models. The model we utilized for this task in shown in Figure 3.5.

#### Asynchronous Task Aware Optimization

Traditionally, MTL models are optimized via obtaining the weighted losses from the multi tasks, while tuning the weights assigned for each loss manually. However, this is a tedious naive approach, which do not let the MTL model reach the optimal minima for the multiple tasks at the same convergence point. Asynchronous Task Aware Optimization [134] is proposed to address this convergence issue by optimizing each survival classification block and grade classification block with the gradients of each block separately.

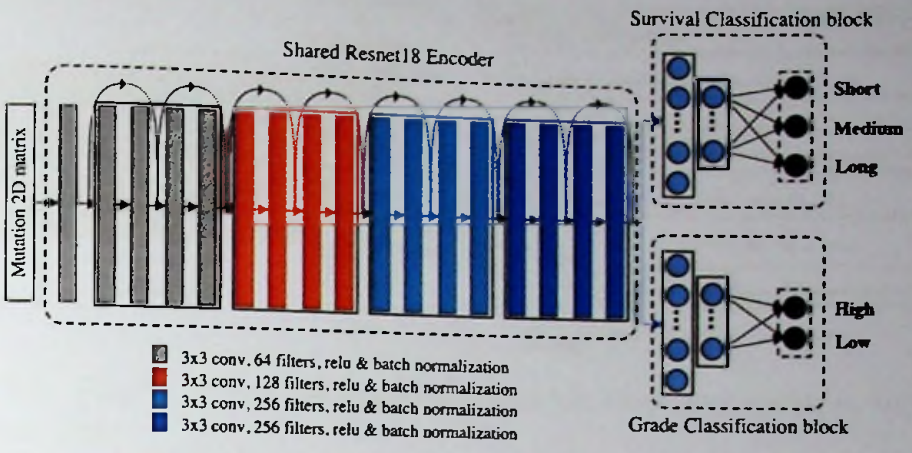



Figure 3.5: MTL model inspired with Resnet18 architecture for survival class and grade prediction.

### GradNorm Asychronous Task Aware Optimization

We introduced the GradNorm Asynchronous Task Aware Optimization, where the survival classification task and the grade classification tasks are optimized independently and then the both tasks are simultaneously optimized to obtain a global convergence point with gradient normalization [135].

Three main steps are followed in our proposed gradnorm-MTL technique. The algorithm is summarized in algorithm 1. First the survival prediction task is optimized while keeping the grade survival prediction block frozen as shown in figure 3.6. Then those optimized blocks, i.e. the shared encoder and the survival prediction blocks are freezed, the grade prediction block is optimized via back propagation as shown in figure 3.7. These two initial steps are shown in Figure ??. Then we incorperated the gradient normalization for the final optimization, where both branches and the shared encoder are unfrozen.

## 3.4 Prognosis with Radiogenomics

### 3.4.1 Dataset

The TCGA-GBM [1] dataset, that consists of WHO grade IV or Glioblastoma cases were used for this radiogenomic analysis. Verhaak et al. [136] had determined

[1] https://www.cancer.gov/tcga

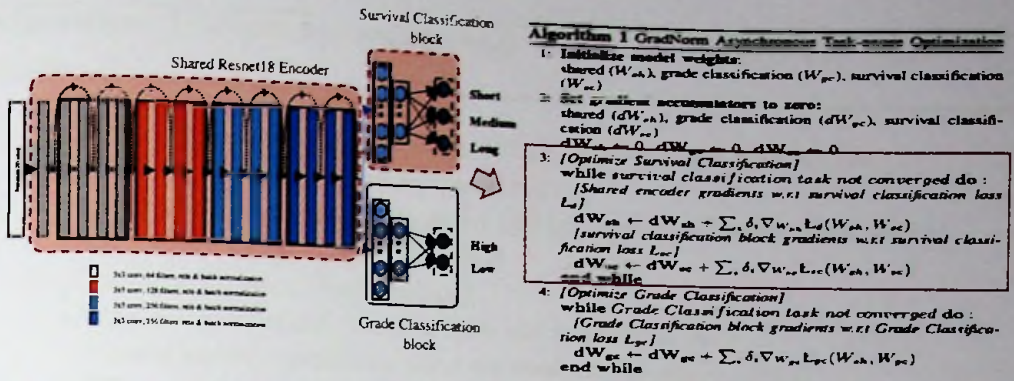



Figure 3.6: Step 1: Freeze the grade classification block and optimize survival prediction task.

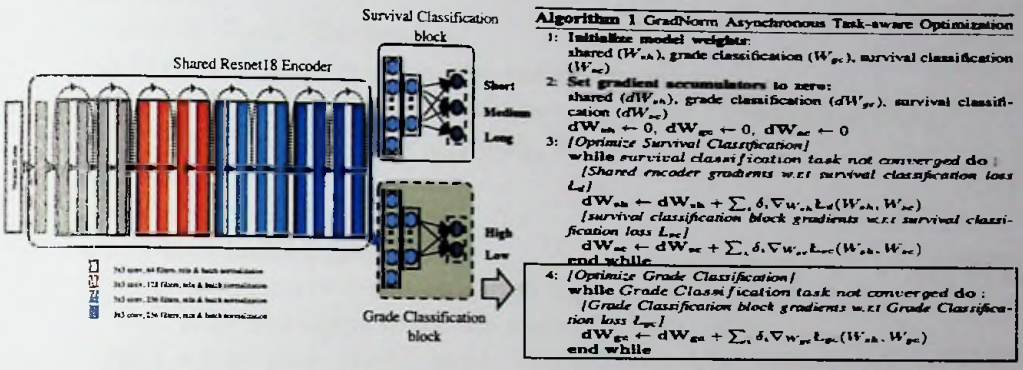



Figure 3.7: Step 2: Freeze the survival classification block and the shared encoder, and optimize grade prediction task.

the 4 subtypes known as Classical, Mesenchymal, Proneural and Neural in their work, which had used 202 TCGA patient cases. We obtained those data from their work, which had the expression profiles and the subtype information. Out of those 202 cases, only 59 cases had the pre operative MRI scans, namely Flair and T1 contrast modalities. Those MRI images were downloaded separately from TCIA, and skull stripped using Matlab. These were in Nifti format and then, they were registered to match to BraTS 2019 dataset using the open-source 3D slicer software [137]. After preprocessing the MRI scans were sized (155*155*240) with the origin located at (0,-239,0). The reason for registering to the BraTS is, we needed a large cohort for the segmentation DL model training, with the manually annotated segmentation ground truths.

Each patient case with the MRI consisted of 1740 gene expression data. Also, the clinical information and the subtype information were also extracted from the

**Algorithm 1:** GradNorm Asynchronous Task-aware Optimization: Algorithms

**Initialize model weights:**
shared ($W_{sh}$), grade classification ($W_{gc}$), survival classification ($W_{sc}$)
**Set gradient accumulators to zero:**
shared ($dW_{sh}$), grade classification ($dW_{gc}$), survival classification ($dW_{sc}$)
$\mathbf{dW_{sh}} \leftarrow 0$, $\mathbf{dW_{gc}} \leftarrow 0$, $\mathbf{dW_{sc}} \leftarrow 0$
*[Optimize Survival Classification]*
**while** *survival classification task not converged* **do** :
*[Shared encoder gradients w.r.t survival classification loss $_d$]*
$\mathbf{dW_{sh}} \leftarrow \mathbf{dW_{sh}} + \sum_i \delta_i \nabla_{W_{sh}} d(W_{sh}, W_{sc})$
*[survival classification block gradients w.r.t survival classification loss $_{sc}$]*
$\mathbf{dW_{sc}} \leftarrow \mathbf{dW_{sc}} + \sum_i \delta_i \nabla_{W_{sc}} sc(W_{sh}, W_{sc})$
**end while**
*[Optimize Grade Classification]*
**while** *Grade Classification task not converged* **do** :
*[Grade Classification block gradients w.r.t Grade Classification loss $_{gc}$]*
$\mathbf{dW_{gc}} \leftarrow \mathbf{dW_{gc}} + \sum_i \delta_i \nabla_{W_{gc}} gc(W_{sh}, W_{gc})$
**end while**
*[Gradient normalization]*
**while** *both tasks improving* **do** :
Pick value for $\alpha > 0$ and pick the weights $\mathcal{W}$ of the final layer of weights which are shared between tasks.
**for** $t = 0$ **to** *max_train_steps* :
**Input** batch $x_i$ to compute $_{sc}(t)$ & $_{gc}(t)$ and $(t) = \sum_i w_i(t)_i(t)$ [standard forward pass]
Compute $G_W^{(i)}(t)$ and $r_i(t)$ $\forall i$
Compute $\overline{G}_W(t)$ by averaging the $G_W^{(i)}(t)$
Compute $L_{\text{grad}} = \sum_i |G_W^{(i)}(t) - \overline{G}_W(t) \times [r_i(t)]^\alpha|_1$
Compute GradNorm gradients $\nabla_{w_i} L_{\text{grad}}$, keeping targets $\overline{G}_W(t) \times [r_i(t)]^\alpha$ constant
Compute standard gradients $\nabla_{\mathcal{W}} L(t)$
Update $w_i(t) \mapsto w_i(t+1)$ using $\nabla_{w_i} L_{\text{grad}}$
Update $\mathcal{W}(t) \mapsto \mathcal{W}(t+1)$ using $\nabla_{\mathcal{W}} L(t)$ [standard backward pass]
Renormalize $w_i(t+1)$ so that $\sum_i w_i(t+1) = T$
**end while**



Verhaak's work.

As we mentioned above, BraTS 2019 dataset was used for training the DL algorith for the segmentation task. For this as the input, only the flair and T1 contrast modalities were used. The ground consisted of three segmented regions known as necrosis; most inner region, enhancement region; region in between necrosis and edema and Edema, outer most region.

### 3.4.2 Segmentation

A pre proposed architecture for BraTS 2019 challenge was used for segmenting the tumor regions. This architecture is a 3D UNet architecture [51], where the decoder consists of 3D attention module [138]. In order to learn more significant and relevant features, spatially squeezing and channel wise excitation is used in this particular attention module. In this module, the channel wise dependencies were captured through 1x1xC convolution of the feature map (with the dimensions of HxWxC), and the spatial feature correlation are captured through max-pooling the feature map. The quality of extracting the region of interest, segmentation is enhanced by changing the model complexity slightly with this module.

155 x 155 x 240 3D MRI volume was given as the input of the segmentation model, to obtain the segmentation feature map. The model was trained with the BraTS manually annotated feature maps. The segmentation of the TCGA dataset was achieved with the trained model giving the FLAIR and T1 Contrast modalities as input.

### 3.4.3 Feature Extraction

After segmenting the tumor, features that can be associated with survival were extracted from each sub region. These features included intensity features, geometrical features and volume related features.

When extracting features, combined regions such as tumor core, the combined region of necrosis and the enhance and the whole tumor region, all the three subregions combined together were considered as well. Thus, we extracted texture and the intensity based features, such as mean intensity, histogram of the oriented gradients, entropy and kurtosis were extracted from the whole tumor, tumor core, enhancement and necrosis regions. The T1 contrast and Flair amplify different subregions in different levels. For an instance, T1 contrast clearly magnify the enhancement region compared to Flair modality. Therefore, when extracting histogram features, features were extracted from both both modalities.

As geometric features, i.e. features that reflect the shapes and size of the

tumor were extracted, were extracted from necrosis, tumor core and whole tumor. Outer subregions were not separately considered as they geometrically function as a single region. Major and minor axis lengths, centroid coordinates, eigen values of the inertia tensor and the inertia tensor are the geometric features we extracted. Bounding box approach with the fractal dimensions was used to extract the volume related features. They were also extracted from the whole tumor, tumor core and the necrosis regions. Altogether, 122 radiomic features were obtained for the analysis.

### 3.4.4 Statistical Analysis

First, we wanted to find the relations between the radiomic features, we extracted from the MRIs and the genomic features. Therefore, the Pearson's correlation was applied to find the correlation between the extracted radiomic features and the gene expression features. In addition, we separately analysed the correlation of each radiomic and the genomic feature with the overall survival of the GBM patients. The same Pearson's correlation was applied for this task. Altogether, both radiomics and genomics, i.e. radiogenomics, association with the survival was analysed using the multivariate and the univariate cox proportional hazard analysis.

Then, we identified the association between the radiomics and the genomic based subtypes of Glioblastoma. This is the most important part of this study, as the subtype is mainly decided using the genomic profiles. If the clinicians can get an idea about the imaging features, that might lead for prior accurate diagnosis without acquiring invasive procedures such as biopsy. Initially, the Kolmogorov-Smirnov test [139] was applied on the radiomic features to verify the normality of each feature distribution. Thus, the features which gave $p < 0.05$ were identified as features distributed normally. Then, Kruskal Wallis non-parametric test [139] was executed on the non normally distributed radiomic features to determine the significantly different radiomics between the 4 GBM subtypes.

In addition, the significantly different subtype pairs were determined by using

the Wilcoxon test [140]. The Bonferroni correction was performed on the multiple comparison tests to avoid the familywise error rate. We used the GraphPad Prism software to visualize the significance of the tests and the feature distribution.

### 3.4.5 Subtype Predictive model

Then, we applied the Machine Learning algorithms to predict the subtypes using the significantly different features. Before feeding to the model, standard normalization was applied. In this work, we only predicted three subtypes, each as a binary classification (subtype/non-subtype). For an example we trained each ML model to predict the subtype is classical or not. The ML algorithms such as Random Forest Classification (RFC), Decision Tree (DT), Linear Kernel function of Support Vector Machine (linear-SVM), radial basis kernel function of Support Vector Machine (r-SVM), polynomial kernel function of Support Vector Machine (p-SVM) and Logistic Regression (LR) were utilized for predicting Classical, Mesenchymal and Proneural subtypes separately with the imaging biomarkers. The parameters used in the above learning models are given in the Table 3.3.

Nevertheless, linear-SVM, r-SVM, p-SVM, DT and RFC were used to predict the subtypes with the differentially expressed genes between the subtypes. This was performed as a 4 class classification (Mesenchymal, Proneural, Neural, and Classical).

Further, four-fold cross-validation was performed to validate the performance of all the subtype predictions. All the experiments were performed with e1071 R package [141] and WEKA software [142].

Table 3.3: Parameters of the subtype prediction learning models

| ML tool | Parameters |
|---|---|
| l-SVM | linear kernel function |
| r-SVM | radial basis function |
| p-SVM | polynomial kernel function |
| DT | depth of the tree: 5 & Number of leaves: |
| RFC | number of trees: 100 |

# Chapter 4

# RESULTS

## 4.1 Prognosis prediction with gene expression profiles

### 4.1.1 System Evaluation - ML approaches & Risk Score Model

#### Overall survival prediction

To compare the performance between the classification ML models we used the metrics such as Accuracy, Recall, Precision and F1 scores. All the ML models were trained on the training split of the each CGGA fold and validated on the corresponding validating split of the CGGA fold. The average over the 4 folds were calculated to measure the performance. The Table 4.1 summarizes the results received from the 4 fold cross validation. We can clearly observe that our proposed BNN algorithms gives the best performance, with an accuracy over 70%, compared to the other commonly used traditional ML algorithms. The second best performing models was the Random Forest Classifier which gave an accuracy over 60%.

Table 4.1: Comparison of Overall Survival Prediction with Machine Learning - 4 fold cross validation on CGGA cohort

| ML methods | Accuracy | Precision | Recall | F1 score |
|---|---|---|---|---|
| DT [74] | 55.25% | 57.50% | 55.25% | 56.00% |
| RFC [84] | 62.25% | 61.25% | 62.25% | 61.00% |
| XGBoost [114] | 56.25% | 62.75% | 56.25% | 58.25% |
| CatBoost [115] | 57.25% | 62.75% | 57.25% | 59.00% |
| SVC [70] | 52.75% | 67.25% | 52.75% | 56.50% |
| **BNN** | **74.50%** | **67.50%** | **73.00%** | 70.25% |

Next, to evaluate the performance of our model on unseen data, we tested the trained model of TCGA dataset. Since we have 4 trained model from each CGGA fold, we tested on each model and considered the average. Even for this,

we could observe that our proposed BNN gives the best performance considering all the metrics compared to the rest of the machine learning algorithms. Out of those ML algorithms, SVM and DT are frequently used in survival prediction with radiomics, as we discussed in the literature review. However, those did not perform well on the gene expression data. Similar to the cross validation results, RFC gave the second best performing results, while the boosting algorithm was the third best. The comparison of the performance of the models on TCGA data is shown in Table 4.2.

According to the state of art literature, the highest reported accuracy is 58.6% with radiomics as the input [48], where a cohort over 100 patients were used for the training. Thus, we could observe that our proposed model works better, quantitatively.

Table 4.2: Comparison of Overall Survival Prediction with Machine Learning - testing on TCGA cohort

| ML methods | Accuracy | Precision | Recall | F1 score |
|---|---|---|---|---|
| DT | 44.50% | 46.00% | 44.50% | 45.00% |
| RFC | 54.25% | 51.25% | 54.25% | 51.00% |
| XGBoost | 51.00% | 52.25% | 51.00% | 51.75% |
| CatBoost | 50.25% | 52.75% | 50.25% | 50.75% |
| SVC | 47.25% | 56.00% | 47.25% | 50.00% |
| BNN | **59.75%** | **57.25%** | 55.25% | 51.00% |



### Prognostic risk model

Risk score models are developed using the coefficients obtained from the univariate cox regression analysis as we mentioned in the literature review. Thus, we also followed the same procedure with our selected 7 gene signature, using the BMA algorithm. TXLNA, STK40, SLC30A7, YTHDF2, SNRNP40, TAF12 and WDR77 genes were used for this model creation.The coefficients we achieved after fitting for a cox regression model is given in the Table 4.3.

As given in the table, all the 7 genes gave a Hazard Ratio over 2, further proving that the genes we selected are highly associated with the survival. Nonethe-

Table 4.3: Univariate Cox Regression analysis on the chosen 7 genes

| Gene | Coef in coxPH | Hazard Ratio (95% CI) | p value |
|---|---|---|---|
| TXLNA | 0.87 | 2.4 (2.1-2.7) | $2.2e^{-36}$ |
| STK40 | 0.77 | 2.2 (1.9-2.4) | $4.8e^{-33}$ |
| SLC30A7 | 0.8 | 2.2 (2-2.5) | $2.6e^{-33}$ |
| YTHDF2 | 0.74 | 2.1 (1.8-2.4) | $1.5e^{-29}$ |
| SNRNP40 | 0.73 | 2.1 (1.8-2.4) | $1.2e^{-31}$ |
| TAF12 | 0.73 | 2.1 (1.8-2.3) | $1.5e^{-34}$ |
| WDR77 | 0.77 | 2.2 (1.9-2.5) | $5.1e^{-32}$ |

less, we considered whether the coefficients are the statistical significant, with p value $< 0.01$. All the coefficients were positive and thus, we can decide that the high expression of these genes have lower survival rate, or poor survival in glioma patients. The risk score formula can be derived as given in Equation 4.1.

$$\begin{aligned}\text{Risk score} = 0.87 * \exp_{TXLNA} + 0.77 * \exp_{STK40} + 0.8 * \exp_{SLC30A7} + 0.74 \\ * \exp_{YTHDF2} + 0.73 * \exp_{SNRNP40} + 0.73 * \exp_{TAF12} + 0.77 * \exp_{WDR77}\end{aligned} \tag{4.1}$$

The above mentioned risk score formula was used to calculate the risk for all our selected glioma patient cases in both CGGA and TCGA. A threshold for high and low risk was set based on the median risk value of the CGGA cohort. Thus, the median risk score was found as 0.6396.

As shown in Fig. 4.1(a), patients with high gene expression levels for the 7 genes clearly had a lower survival, further implying the high risk they hold. On the other hand, the lower risk people showed a lower expression level for all the 7 genes, thus proving to have a longer survival.

This prognostic risk model was validated on the TCGA cohort by obtaining the risk score using the Equation (4.1). The same threshold was used to divide the patients into high and low risk groups. Consequently, the aforementioned relationship could also be seen when we observe the Fig 4.1 (b) obtained for the TCGA cohort. Most of the patients with a highly expressed 7 genes also had a low survival and a high risk.

To make this more clear, we visualized a plot,Fig. 4.2 for each risk categories separtely for both data cohorts. We could observe that the high risk people have

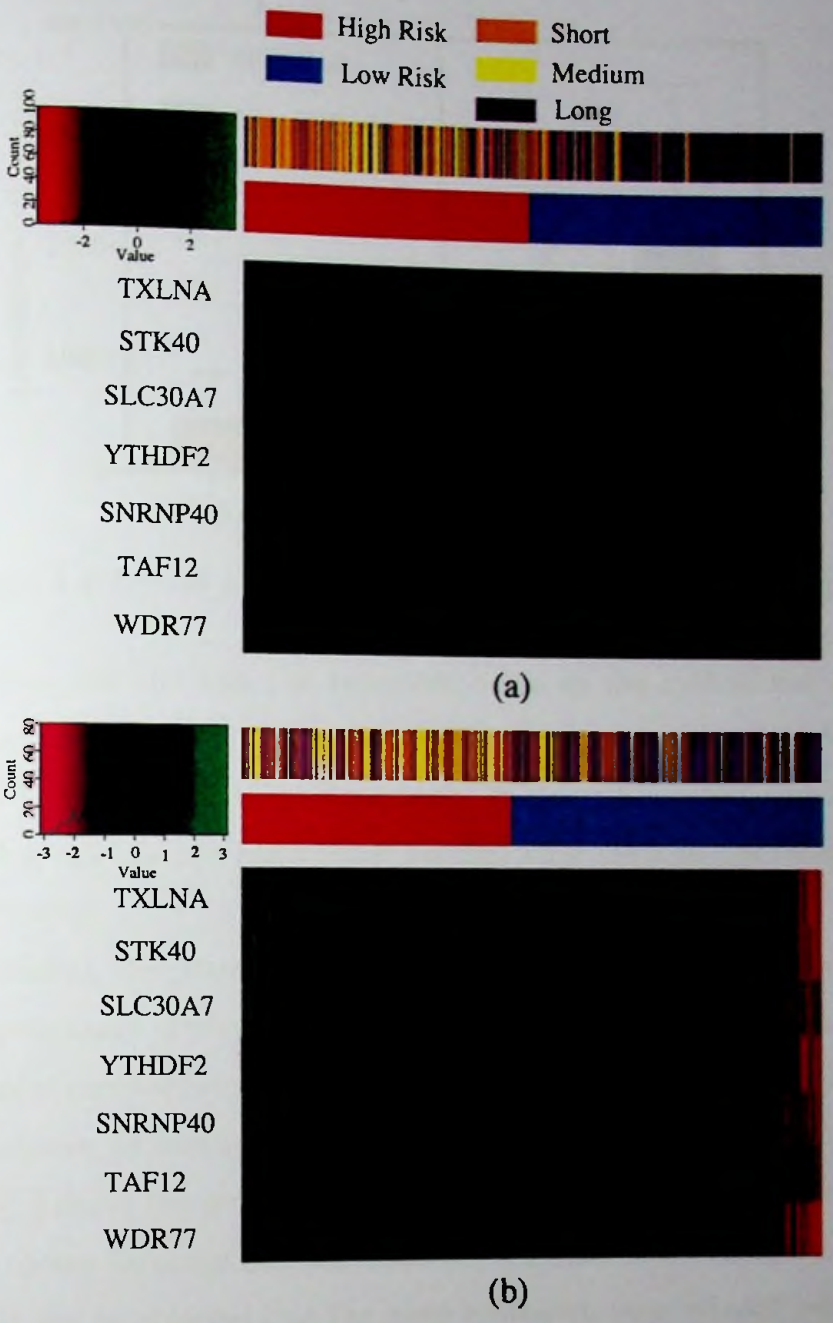



Figure 4.1: Heat map for the (a) CGGA cohort (b) validation TCGA cohort with proposed gene signature.

low span of overall survival in days, compared to the low risk category shown in blue for both datasets.

Quantitatively, the CGGA cohort has a mean of 573.2658 days and 2244.361 days, with a standard deviation of 662.5993 days and 1385.437 days respectively for high risk and low risk categories. The testing cohort, TCGA, which was

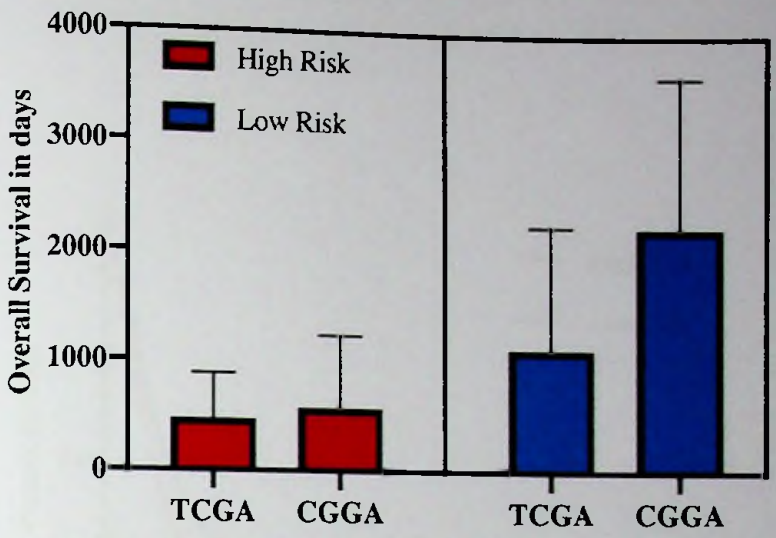


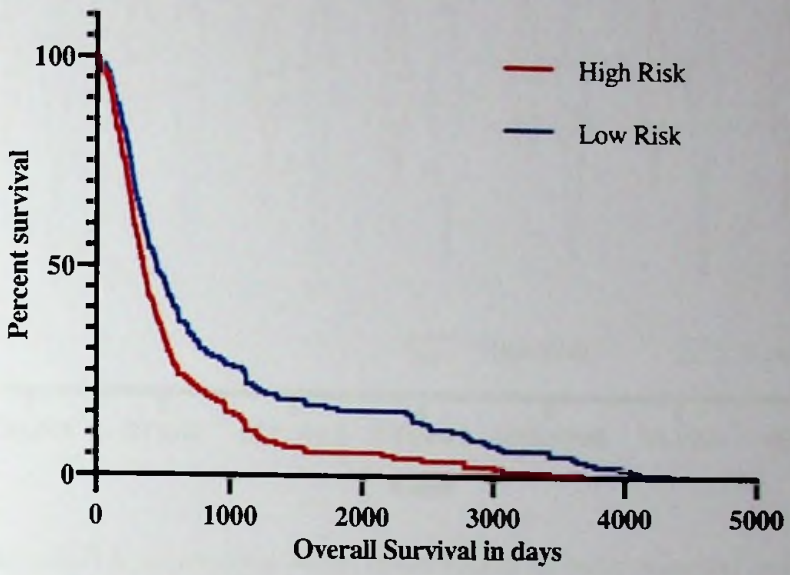



Figure 4.2: Overall survival distribution of high and low risk groups.

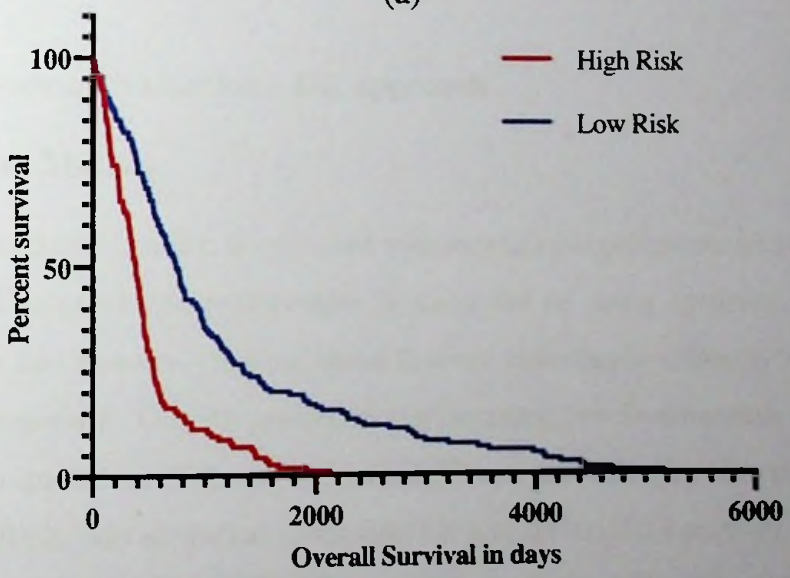
divided in to low and high risk categories based on the median risk score of CGGA, gave a mean of 468.1624 days and 1124.43 days with a standard deviation of 412.3076 days and 1131.1 days respectively. In contrast, we could clearly notice that the low risk category has a higher survival while the high risk category has a lower survival.

Nevertheless, the plotted Kaplan-Meier graphs, shown in Fig. 4.3 verify the reduced percentage of survival in high risk category with respect to the higher percentage of survival in low risk category. This was evident for both CGGA and TCGA datasets, as shown in subfigures (a) and (b).

Fig. 4.4 shows the gene expression distribution of the the most prominent genes we choose for prognostic risk model development, in high risk and low risk groups. It can be observed that the mean expression value of each gene in the low risk group is lower than the mean expression value of the corresponding gene in the high risk group.




(a)




(b)

Figure 4.3: Kaplan-Meier curves obtained for the high risk and low risk groups. (a) CGGA (b) TCGA dataset.

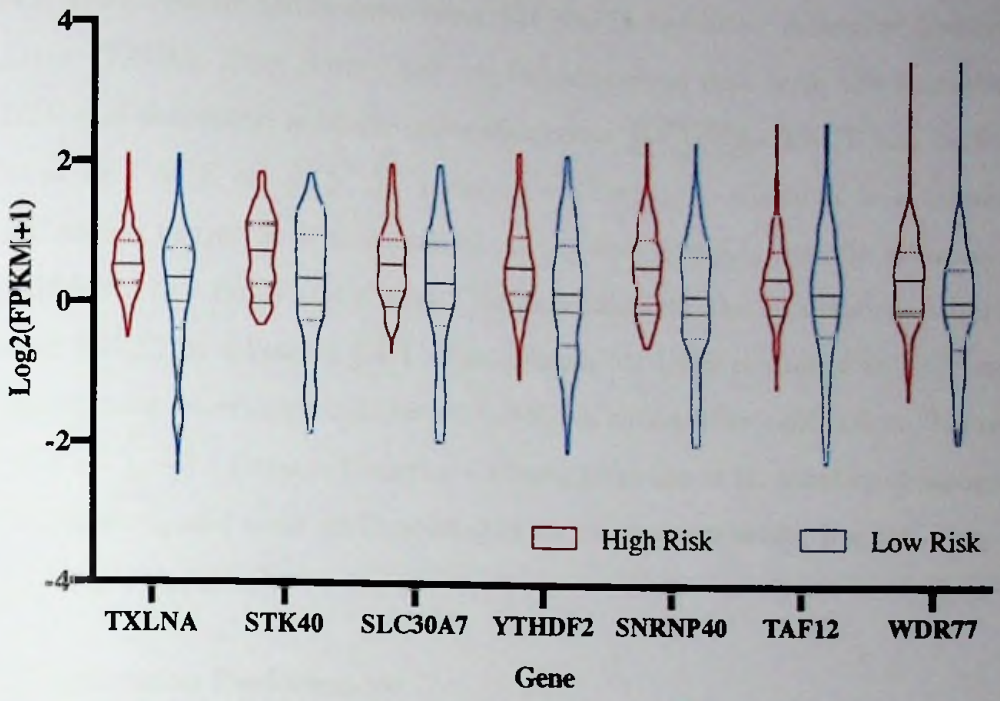



Figure 4.4: mRNA expression value distribution of each selected genes for the high and low risk groups.

### 4.1.2 System Evaluation - DL approach

#### Evaluation Metrics

The proposed loss function is evaluated with extensive experiments and evaluation metrics. The prediction performance is measured by using accuracy, precision, sensitivity and f1-score. Among these f1-score reflects the capacity of individual class accuracy. Despite prediction performance, we demonstrate the model calibration and uncertainty on our normLSF over the CE. For this purpose, we utilize multiple miscalibration quantification and uncertainty metrics. Expected Calibration Error (ECE) [143, 144] is used as the base calibration metric. ECE computes the discrepancy between accuracy and confidence probability. It can be formulate as, $ECE = \sum_{b=1}^{B} \frac{n_{[b]}}{N} \mid Acc_{[b]} - Conf_{[b]} \mid$, where $B$, $N$, $n_[b]$, $Acc_{[b]}$ and $Conf_{[b]}$ are the total number of bins, total number of samples, and number of samples, accuracy and confidences in $b^{th}$ bin. However, it is indicated that ECE is sensitive to bin number and unable to present class-wise model calibra-

tion [145]. Static Calibration Error (SCE), Thresholded Adaptive Calibration Error (TACE), Brier Score (BS) are introduced to deal with the limitation of ECE and determine accurate calibration error [145, 146]. TACE can be formulated as $TACE = \frac{1}{KR}\sum_{k=1}^{K}\sum_{r=1}^{R} \mid Acc_{[r,c]} - Conf_{[r,c]} \mid$, where $R$ is a number of calibration ranges with threshold t, $Acc_{[r,c]}$ and $Conf_{[r,c]}$ are the accuracy and confidence in r range and c class. We also calculate the uncertainty calibration error (UCE) by following [147]. The formula for UCE is similar to ECE except determining uncertainty calibration instead of probability calibration. Therefore, $UCE = \sum_{b=1}^{B} \frac{n_{[b]}}{N} \mid Err_{[b]} - Uncert_{[b]} \mid$ where, total bin is B, number of samples in $b^{th}$ bin is $n_{[b]}$ and error and uncertainty in the corresponding bin are $Err_{[b]}$ and $Uncert_{[b]}$ respectively.

### Classification Performance

Table 4.4: Performance study of state-of-the-art CNN models trained with cross-entropy and proposed loss for classifying of 4 cross-validated CGGA genomic data samples. CE, normLSF, Acc., Prec., Sens. represents Cross-Entropy, Normalized Label Smoothing Focal, Accuracy, Precision and Sensitivity respectively.

| Models | Cross-Entropy (CE) | | | | normLSF | | | |
|---|---|---|---|---|---|---|---|---|
| | **Acc.** | **Prec.** | **Sen.** | **F1** | **Acc.** | **Prec.** | **Sen.** | **F1** |
| ResNet18 [121] | **0.682** | 0.661 | 0.728 | 0.678 | 0.672 | **0.663** | **0.765** | **0.696** |
| VGG16 [42] | 0.647 | 0.639 | 0.692 | 0.625 | **0.653** | **0.623** | **0.724** | **0.660** |
| GoogleNet [122] | 0.634 | 0.556 | 0.652 | 0.589 | **0.676** | **0.711** | **0.744** | **0.664** |
| MixConv [123] | 0.663 | 0.618 | 0.666 | 0.629 | **0.682** | **0.665** | **0.732** | **0.667** |
| ResNest50 [124] | 0.679 | 0.661 | 0.702 | 0.666 | **0.695** | **0.683** | **0.721** | **0.675** |
| ShuffleNet [125] | 0.666 | 0.624 | 0.666 | 0.630 | **0.673** | **0.655** | **0.695** | **0.660** |

The classification performance of our proposed loss function is evaluated with the metrics of accuracy, precision, sensitivity, and f1 score. As both datasets, CGGA, and TCGA, consist of imbalanced classes; therefore, f1-score reflects the generalizability of the model. Table 4.4 represents the mean comprehensive evaluation of various state-of-the-art classification models on normLSF over CE loss with the CGGA dataset. It is clear that normLSF significantly outperform CE in all the evaluation metrics. In particular, there is around 4% f1-score improvement, which proves the effectiveness of normLSF to deal with the class

Table 4.5: Average classification metrics for validation-set obtained with different loss functions on a 4-fold CGGA dataset. The values manifested in bold are best per column. CE, FL, LSCE, LSF, normLSF, Acc., Prec., Sens. represents Cross-Entropy, Focal Loss, Label Smoothing Cross-Entropy, Label Smoothing Focal, Normalized Label Smoothing Focal, Accuracy, Precision and Sensitivity respectively.

| Loss | | | | | Acc. | Prec. | Sens. | F1. |
|---|---|---|---|---|---|---|---|---|
| CE | FL [127] | LSCE [129] | LSF | normLSF | | | | |
| × | × | × | × | ✓ | 0.67 | **0.66** | **0.76** | **0.69** |
| × | × | × | ✓ | × | 0.68 | 0.66 | 0.73 | 0.68 |
| × | × | ✓ | × | × | 0.67 | 0.63 | 0.72 | 0.66 |
| × | ✓ | × | × | × | **0.70** | 0.67 | 0.70 | 0.66 |
| ✓ | × | × | × | × | 0.64 | 0.61 | 0.73 | 0.66 |

Table 4.6: 4-fold cross evaluation metrics of ResNet18 model trained on CGGA Genomics dataset using ours loss.

| #Fold | Accuracy | Precision | Sensitivity | F1-Score |
|---|---|---|---|---|
| 1 | 0.6329 | 0.6471 | 0.7143 | 0.6772 |
| 2 | 0.7089 | 0.6958 | 0.8000 | 0.7381 |
| 3 | 0.6962 | 0.6847 | 0.7971 | 0.7089 |
| 4 | 0.6538 | 0.6257 | 0.7500 | 0.6628 |
| Mean | **0.6729** | **0.6633** | **0.7653** | **0.6967** |

skewness issue. A fold wise performance scores of ResNet18 are depicted in Table 4.6. We choose ResNet18 for further result analysis as it achieves the highest accuracy among all the classification models. Table 4.5 shows the ResNet18 performance with different state-of-the-art loss functions and our proposed LSF and normLSF loss functions. A notable enhancement of the performance is shown with our functions. Specifically, accuracy and f1-score are increased by around [1.5%, 3%] and [4% 5%] with LSF and normLSF, respectively. We also observe that LS with focal loss (LSF) is more effective than LS with CE (LSCE), where LSCE is not helping in the class-skewness problem. Figure 4.5

We also conduct experiments on the TCGA dataset to confirm the results of our proposed loss function. Fig 4.5 illustrates the evaluation metrics of normLSF over CE for ResNet18 with TCGA dataset. Similar to CGGA, normLSF achieves the best results in TCGA with 2-3% performance improvement, while outperforming CE loss in terms of model prediction metrics.

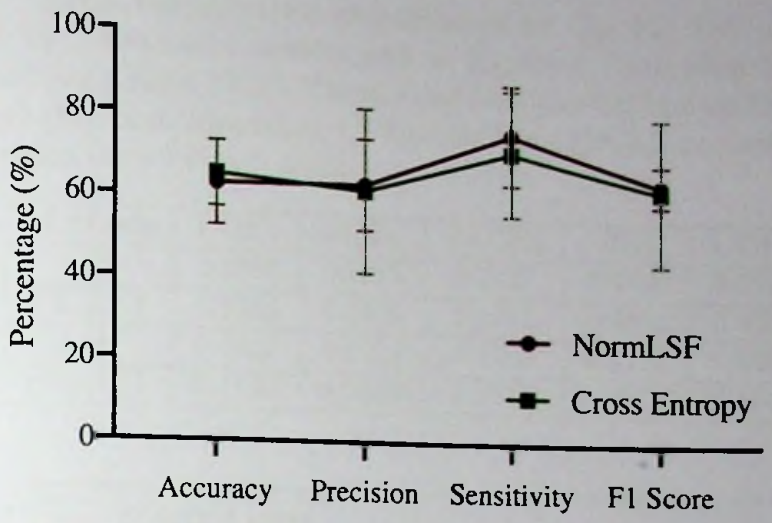



Figure 4.5: Comparative classification performance analysis for model trained with state-of-the-art cross entropy and proposed cost function on TCGA dataset.

### Model Calibration and Uncertainty

In addition to classification performance, we also determine the reliability of prediction using different miscalibration and uncertainty metrics. For example, Expected Calibration Error (ECE), Static Calibration Error (SCE), Thresholded Adaptive Calibration Error (TACE), Brier Score (BS), and Uncertainty Calibration Error (UCE) metrics are utilized to measure the model calibration and uncertainty. Table 4.7 demonstrates the miscalibration quantification for CE, FL, our LSF and normLSF using ResNet18 on CGGA dataset. It is clear that normLSF produces better calibration and outperforms other loss functions in most of the metrics. The model with LSF obtains better uncertainty than normLSF. However, LSF and normLSF represent superior model calibration comparing to CE and FL loss functions. A reliability diagram of these loss functions is illustrated in Figure 4.6. The diagonal dash line is considered as the perfect calibrated model where the confidence score vs. accuracy graph is expected to be aligned, and the reliability curve closer to the diagonal line is presented less calibration error. In the figure, normLSF curve is more closer to the perfect calibration comparing to other loss functions. Thus, CE and FL show higher miscalibration and far deviated from the dash line.

Table 4.7: Model miscalibration quantification for CE, FL, LSF and proposed normLSF loss. Evaluation metrics such as Expected Calibration Error (ECE), Static Calibration Error (SCE), Thresholded Adaptive Calibration Error (TACE), Brier Score (BS) and Uncertainty Calibration Error (UCE) are used to calculate the calibration error for each model.

| Methods | ECE↓ | SCE↓ | TACE↓ | BS↓ | UCE↓ |
|---|---|---|---|---|---|
| CE | 0.1998 | 0.1613 | 0.2127 | 0.4820 | 0.5972 |
| FL [127] | 0.1933 | 0.1476 | 0.1911 | 0.4666 | 0.6030 |
| LSF | 0.1452 | 0.1223 | 0.1369 | 0.4177 | **0.1464** |
| normLSF | **0.1052** | **0.0820** | **0.1306** | **0.4070** | 0.4648 |

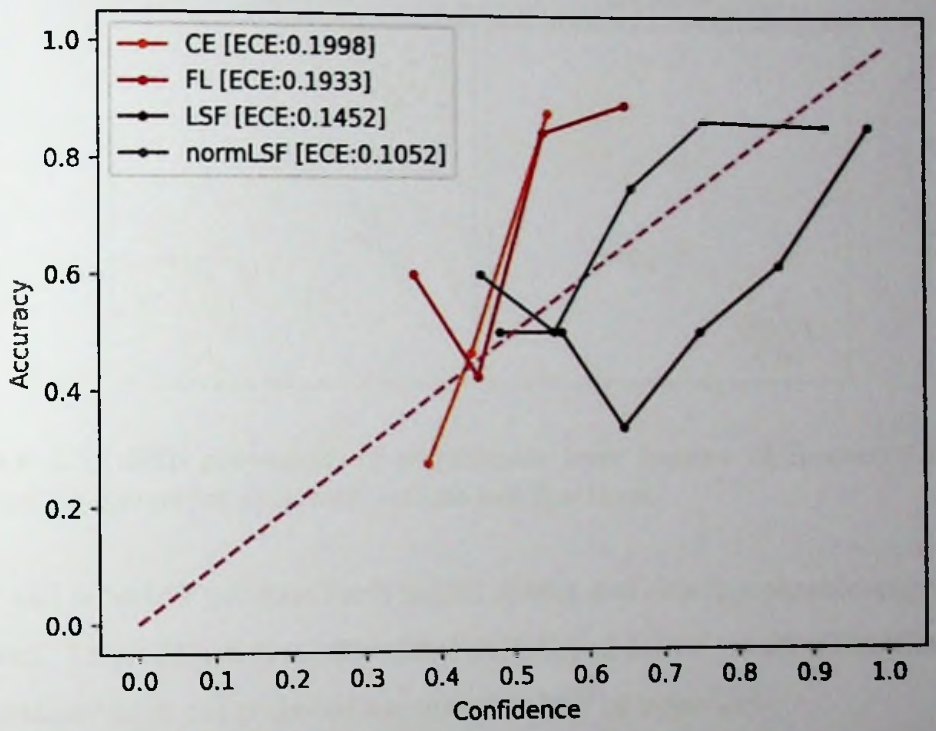



Figure 4.6: Reliability graph for normLSF comparing with CE, FL and LSF. The curve for normLSF is much closer to the perfect reliability cure (dash line).

Figure 4.7 shows the t-SNE (t-distributed Stochastic Neighbor Embedding) plot of feature representation from the ResNet18 penultimate layer trained on CE, FL, LSF, and normLSF. Our LSF and normLSF demonstrate better feature representation with tighter cluster for same class features over CE and FL. We follow [130] to plot the t-SNE from the feature map of the penultimate layer. We observe that The projections are into a broad cluster with CE and FL. Whereas,

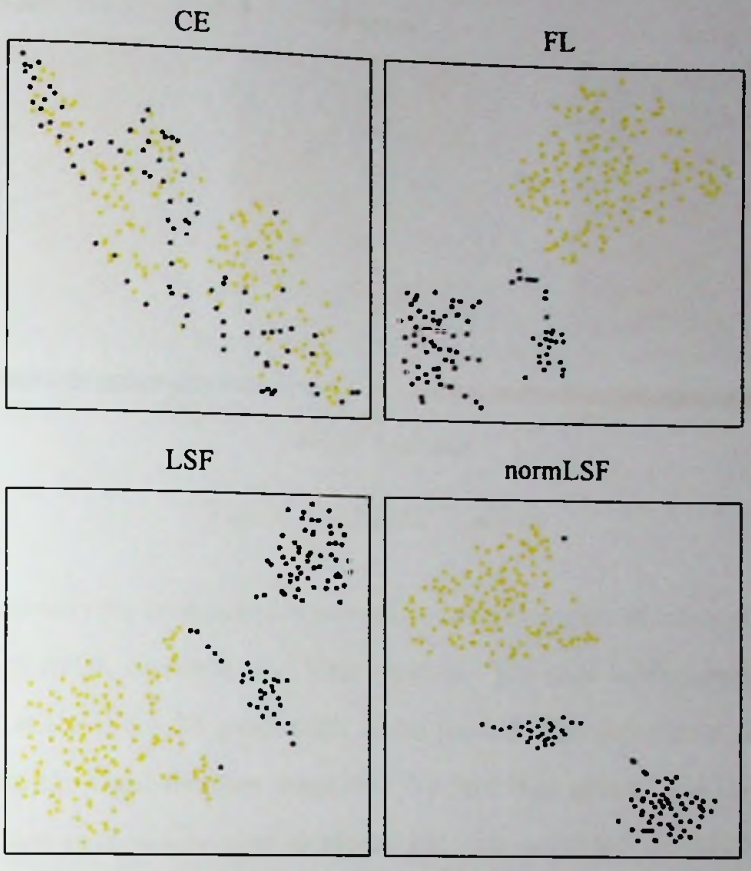





Figure 4.7: tSNE projections of penultimate layer features of ResNet18 model trained on Genomics data with various loss functions.

LSF and normLSF produce much tighter cluster and clearly separable among the classes. These demonstrate the visual evidence of better model calibration and uncertainty with our proposed loss function LSF or normLSF.

### Explainability Analysis

To visualize the model explainability, we utilize SHAP value analysis [148, 149] and integrated gradient algorithm [150] and highlight the relevant and irrelevant genes in three different samples belonging to different classes, shown in Figure 4.8. Here, red pixels intensify the output probability of true class while bluish alleviate the prediction probability. The plot obtains by using Gradient Explainer[1] with 0.5 as the smoothing factor is displayed below. Our model focuses on a particular set

[1] https://shap.readthedocs.io/en/latest/examples.html

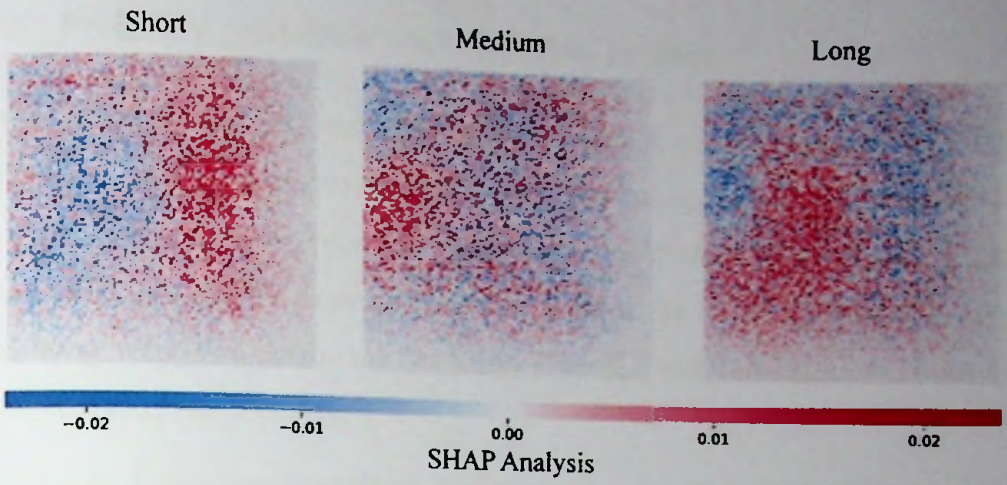



Figure 4.8: SHAP analysis

of genes for classifying true samples correctly, and the region of interest varies with true classes of short, medium, and long survival. The plot in Figure 4.9 indicates the contrast scale of the 20 genes with 10 on the left side signifies positive SHAP values and on the right denotes negative. We find that genes LGALS3, ARNTL, and VPS11 are profoundly crucial biomarkers for accurate classification of the small, medium, and long survival classes, respectively. Likewise, genes TTCS, SHC1, GLE1 are the least significant in the model prediction.

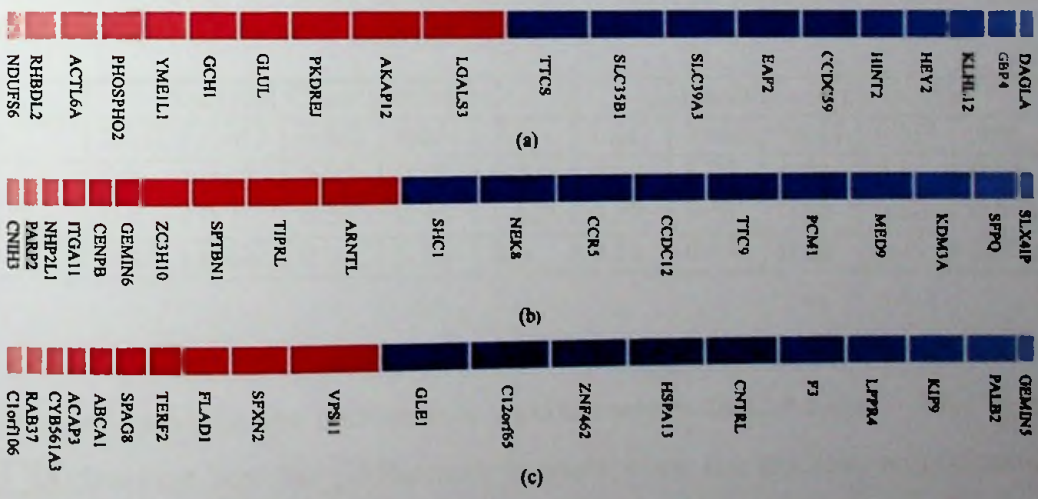



Figure 4.9: The most crucial clinical genes responsible for true and false prediction in red and blue, respectively of all three categories.

## 4.2 Prognosis prediction with mutation profiles

As we mentioned in methods section, we have trained a multi task learning model, to predict the glioma grade and the survival class using the mutation profiles. For this we propose a novel optimization algorithm, GAP-ATL. We utilized the label smoothing loss and the cross entropy loss both in the initial optimization steps. The cross validation results are briefed in Table 4.8 and Table 4.9 for CGGA and TCGA cohorts respectively.

Table 4.8: Prediction of subtype and grade as a multi task classification using mutation profiles on CGGA dataset. Step 1. freeze grade prediction branch(Label Smoothing) Step 2. freeze shared encoder + survival prediction branch (Label Smoothing) Step 3. both unfreeze - GradNorm

| Step | Survival Class prediction | | | | | Glioma grade prediction | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | fold1 | fold2 | fold3 | fold4 | Avg | fold1 | fold2 | fold3 | fold4 | Avg |
| 1 | 0.69 | 0.71 | 0.71 | 0.72 | 0.708 | 0.65 | 0.32 | 0.36 | 0.5 | 0.458 |
| 2 | 0.69 | 0.71 | 0.71 | 0.72 | 0.708 | 0.65 | 0.68 | 0.64 | 0.66 | 0.658 |
| 3 | 0.88 | 0.85 | 0.86 | 0.81 | 0.85 | 0.51 | 0.45 | 0.5 | 0.49 | 0.488 |

Table 4.9: Prediction of subtype and grade as a multi task classification using mutation profiles on TCGA dataset. Step 1. freeze grade prediction branch(Label Smoothing) Step 2. freeze shared encoder + survival prediction branch (Label Smoothing) Step 3. both unfreeze - GradNorm

| Step | Survival Class prediction | | | | | Glioma grade prediction | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | fold1 | fold2 | fold3 | fold4 | Avg | fold1 | fold2 | fold3 | fold4 | Avg |
| 1 | 0.6 | 0.47 | 0.59 | 0.45 | 0.528 | 0.29 | 0.73 | 0.34 | 0.8 | 0.54 |
| 2 | 0.6 | 0.47 | 0.59 | 0.45 | 0.528 | 0.79 | 0.88 | 0.85 | 0.85 | 0.843 |
| 3 | 0.47 | 0.48 | 0.5 | 0.55 | 0.5 | 0.81 | 0.82 | 0.88 | 0.79 | 0.825 |

The final optimization step, have been trained with two different approaches in order to verfiy the performance enhancement with Gradnorm. The figure 4.10 visualizes how the performance increase when the gradient normalization is used in the final optimization without using the traditional backpropagation optimization.

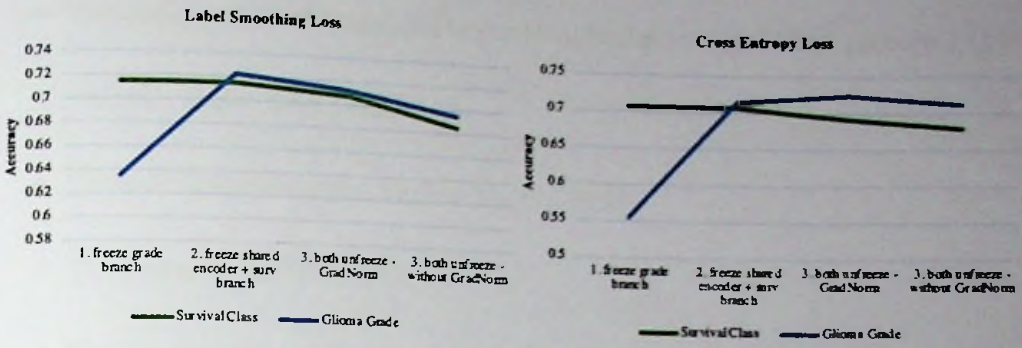



Figure 4.10: Performance comparison between using Label smoothing and Cross entropy loss with and without gradnorm.

## 4.3 Prognosis Analysis with Radiogenomics

### 4.3.1 Correlation between radiomics, genomics and overall survival

$p < 0.05$ and correlation $> 0.3$ is considered as the condition to the features in low positively correlation. This condition was satisfied by 2 imaging features, out of the 119 features extracted from MRI. Both features are kurtosis features extracted from enhancement region using flair and T1 contrast MRI modalities. Similarly, cox proportional hazard analysis gave a negative coefficients significantly ($p < 0.05$). Therefore, we suppose that the higher kurtosis values extracted from the enhancement region implies a large survival span of glioma patients.

In addition, the connections between the overall survival and gene expression was studied. For this 1023 genes were used. Out of them, 25 genes showed a significant ($p < 0.05$) correlation with the survival of glioma patients.

3 of these genes, Endothelin Receptor Type A (EDNRA), Collagen Type III Alpha 1 Chain (COL3A1) and Olfactomedin-like protein 3 (OLFML3) showed a low positive correlation. The other genes which we have listed below gave a low negative correlation.

- Pirin (PIR)
- Secretogranin-2 (SCG2)
- Insulin Like Growth Factor Binding Protein 3 (IGFBP3)

- Mucosa-associated lymphoid tissue lymphoma translocation protein 1 (MALT1)
- Tubulin alpha-4A (TUBA4A)
- Growth Differentiation Factor 15 (GDF15)
- Tetraspanin-13 (TSPAN13)
- Neural precursor cell expressed developmentally downregulated gene 4-like (NEDD4L)
- Endoplasmic Reticulum Aminopeptidase 2 (LRAP)
- Phospholipid Scramblase 1 (PLSCR1)
- Ankyrin Repeat And MYND Domain Containing 2 (ANKMY2)
- IQ Motif Containing G (IQCG)
- Rho Family GTPase 3 (RND3)
- Radical S-Adenosyl Methionine Domain Containing 2 (RSAD2)
- Pyruvate Dehydrogenase Kinase 1 (PDK1)
- Regucalcin (RGN)
- Dihydrouridine Synthase 4 Like (DUS4L)
- Oligoadenylate Synthetase 1 (OAS1)
- FAM3 Metabolism Regulating Signaling Molecule C (FAM3C)
- Membrane Palmitoylated Protein 6 (MPP6)

Our observation further proved, as they similarly gave a Hazard ration $> 1$ with $p < 0.05$, when we performed the univariate cox proportional analysis.

After analysing each genomic and radiomic features' correlation with survival, we assessed the correlation between the genomic and radiomic features as well. We observed that, TIMP Metallopeptidase Inhibitor 4 (TIMP4) gene's expression

shows a moderate positive correlation (correlation > 0.5) with the fractal dimension of the necrosis region. Correspondingly, the extent of the necrosis region had a moderate positive correlation with the acyl-CoA oxidase 2 (ACOX2) gene expression level.

### 4.3.2 Imaging biomarkers associated with molecular subtypes

As we mentioned in the methods section we used the Kolmogorv-Smirnov test to verify the normality of the imaging features we extracted. Out of 119 imaging features, 38 were normally distributed significantly ($p < 0.05$).

Kruskal Wallis test was then applied on non normal features as it is a non parametric test, to identify the features that are different between the 4 Glioblastoma subtypes. We could notice tht 12 features are significantly different between 4 subtypes. These features are the fractal dimension features extracted from the whole tumor, tumor core and the necrosis. All these features are given in the Table 4.10, with the test statistic and p value obtained.

Table 4.10: Kruskal Wallis statistics for radiomic features ($p < 0.05$) between 4 subtypes.

| Feature | test statistic | p value |
|---|---|---|
| tumor core bounding box 1 | 8.8889 | 0.0308 |
| tumor core fractal dimension 3 | 7.9312 | 0.0474 |
| whole tumor fractal dimension 4 | 7.8459 | 0.0493 |
| whole tumor fractal dimension 5 | 7.9820 | 0.0463 |
| kurtosis of necrosis - Flair | 9.197 | 0.0267 |
| bounding box - 1 of necrosis | 9.263 | 0.0259 |
| bounding box - 3 necrosis | 9.0134 | 0.0291 |
| fractal dimension 1 of necrosis | 8.887 | 0.0308 |
| fractal dimension 2 of necrosis | 9.1918 | 0.0268 |
| fractal dimension 3 of necrosis | 10.418 | 0.0153 |
| fractal dimension 4 of necrosis | 9.1716 | 0.027 |
| fractal dimension 5 of necrosis | 8.937 | 0.0301 |



These observations confirm that these radiomic features are significantly different between subtypes, which can be between 2 or more. Thus, we wanted to verify between which subtypes they are different. To analyse this we used the Wilcoxon test between each subtype pair.

The fractal dimension of the necrosis, tumor core and whole tumor region showed a significant difference ($p < 0.0001$) between Proneural and Classical subtypes and also a significant difference ($p < 0.0001$) between Proneural and Mesenchymal subtypes. Moreover, the fractal dimension of the whole tumor showed a significant difference between Classical and Mesenchymal subtypes. The associations and the distributions of the fractal dimensions are shown in the Fig 4.11

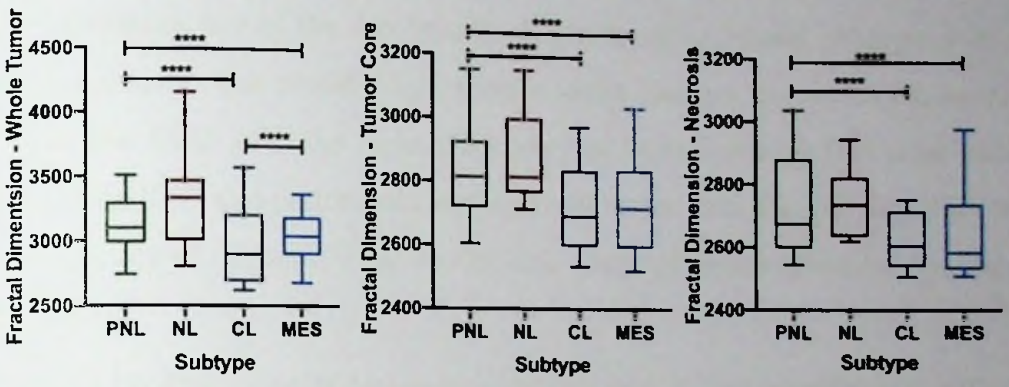



Figure 4.11: The comparison of the fractal dimensions of whole tumor, tumor core and necrosis regions between 4 subtypes of Glioblastoma.

### 4.3.3 Subtype Prediction

As we mentioned above, we identified 12 imaging features using the Kruskal-Wallis test that are significantly different between the 4 subtypes. We then used these features to predict the subtype using machine learning algorithms. First, the Mesenchymal subtype (Mesenchymal - 23 cases, Non-Mesenchymal - 36) was predicted with ML tools and SVM outperformed the other algorithms with an accuracy of 62.7119%, along with 61% precision and 62.7% recall. Next, the Classical subtype (Classical - 12, Non-Classical - 47) is predicted and the SVM outperformed other 3 algorithms with 4 fold validation accuracy of 85.3%. Nevertheless, Prediction of Proneural subtype (Proneural - 16, Non-Proneural - 43) was performed obtaining 4 fold accuracy of 81.82% accuracy with SVM. The performance comparison of the binary classification of each subtype are given in the Table 4.11

Table 4.11: Prediction of subtypes as a binary classification using Radiomics

| ML | Classical | | | Proneural | | | Mesenchymal | | |
|---|---|---|---|---|---|---|---|---|---|
| | Acc | Prec | Recall | Acc | Prec | Recall | Acc | Prec | Recall |
| SVM | 85.3% | 91.4% | 88.8% | 81.8% | 88.8% | 88.8% | 82.0% | 79.2% | 82.0% |
| DT | 83.1% | 83.1% | 83.1% | 66.1% | 63.3% | 66.1% | 57.6% | 58.7% | 57.6% |
| RFC | 79.7% | 77.3% | 79.7% | 69.4% | 62.7% | 69.5% | 61.0% | 60.2% | 61.0% |
| LR | 77.9% | 80.1% | 78.0% | 64.4% | 65.1% | 64.4% | 57.6% | 56.7% | 57.6% |

The subtype prediction with genomics was executed as a generalized method for predicting any of the 4 subtypes. However, since Neural subtype does not exist anymore, this model might require slight changes in current exploration. Thus, the SVM gave the highest accuracy of 94.91% with 4 fold cross validation. The RFC also performed comparatively better than the DT algorithm with an accuracy of 89.93%. This verifies that usage of gene expression despite the acquition

Table 4.12: Prediction of Molecular subtypes as a 4 class classification with Genomics

| ML method | Accuracy | Precision | Recall |
|---|---|---|---|
| SVM | 94.91% | 95.4% | 94.9% |
| DT | 64.4% | 63.3% | 64.4% |
| RFC | 89.83% | 90.8% | 89.8% |

# Chapter 5

# DISCUSSION

## 5.1 Comparison with existing studies

The state of art survival prediction tools were developed mainly using the imaging features, i.e. radiomics. A well known fact regarding the prognosis of gliomas and the occurence of gliomas, i.e. the gliomagenesis, mainly rely on the alterations that occur in the molecular level, which cannot be identified by obtaining images. More specifically the heterogeneity, or the large variations in those gene profiles, makes it quite complex to decide the behavior of the cancer. However, assessing large data cohorts is not also efficient as we do not have large number of cases/instances for training too. Therefore, it is necessary to identify the most vital genes associated with the prognosis. Many tools and algorithms were developed recently by identifying and utilzing the survival associated genes. In this work, we identify 7 genes, that the expression levels depicts the survival of glioma patients.

These selected gene signature, was used to develop a Bayesian Neural Network to predict the survival. Our experiments demonstrate the this model has the ability to predict the survival of the unseen data more accurately compared to the state of art tools. The state of art tools, we compared our proposed algorithm, were mostly developed using the features extracted from images. The MICCAI BraTS challenge is one of the leading competitions that focus on segmenting and survival prediction of gliomas. The all time best performing model [48] proposed for BraTS gave 68% accuracy with the radiomics. With our work, we proposed an outperforming approach, with over 70% accuracy with gene expression data. In addition to that, if someone wants know the risk with the expression profiles, we propose a risk based method for prognosis prediction as well. The high throughput genomic profiles, unfold many hidden factors about gliomas, than the

radiomics leading for higher accuracies.

We identified 7 genes named, SNRNP40, STK40, TAF12, TXLNA, SLC20A7, WDR77 and YTHDF2, as the most survival related genes through their expression. Out of those, TAF12 was recognized as a gene that has a connection with gliomas in past research work. As we mentioned in the introduction, WHO grade II and III gliomas are said to have mutated IDH1 and IDH2 genes. Ren et al [151] points out that the mutation in IDH1 can downregulate the expression of the TAF12 gene. This finding further on par with our results, which we observed that the high risk patients to have low expressed TAF12 genes, as visualized in Fig. 4.1. Nonetheless, we were able to identify 6 other genes, that have close associations with the survival and can be used for prognostic tool development.

Most of the studies related to genomic analysis of gliomas face obstacles as they were acquired using different protocols that ultimately cause inconsistencies the genomic profiles. Therefore, researchers are tend to rely on a single cohort in their studies. This can happen because of the technological drawbacks in distinct high throughput gene acquisition platforms. Other than that, the profiles can also vary due to the cross hybridization of using probes, probes' annotations and redundancies [152]. Consequently this limitation was also present in our study. Therefore, we could observe a lower performance when testing on an external cohort. For an example, we used CGGA for training and TCGA for testing. We could clearly observe a performance difference in two cohorts. However, when we consider the risk score model, as visualized in fig. 4.1 (b), somewhat comparatively success could be seen, as it has separated the high risk and low risk patients similar to the survival to a certain extend. Thus, we can conclude that our prognostic model is a promising approach compared to the machine learning techniques.

Having not sufficient data for training a learning algorithm is also a drawback for not getting high performance with machine learning. This is a common scenario in many machine learning based glioma prognosis models. For an instance, training a glioma segmentation model with the 3D modalities makes the number of training instances limited, as we have to give the complete 3D volume

as the input of the 3D DL architectures. As a result, many DL researchers are tend to explore 2D architectures for segmenting, as we can create more training instances from a single 3D modality, by giving slice by slice as the input for the DL algorithm.

Our prognostic risk model, is also unique due to its applicability on any glioma type. Previous studies have focused on developing such models, to specific glioma types such as for Glioblastoma or lower grade gliomas. Accordingly, our model can be employed without determining the glioma grade. In addition, this model do not require large number of genes' expression profiles, as this only require 7 genes. Therefore, the complexity is low compared to the complicated models with large number of profiles used in risk estimation.

As stated in state of art Machine Learning literature, the probabilistic programming languages have the ability to learn using a lower number of instances [153]. Since we have this issue on data deficiency with our dataset, this algorithm might have worked better than the other traditional algorithms. In the future, with more data, this can be extended by integrating deep learning and the probabilistic programming in depth as in [116]. We only applied shallow bayesian neural networks, as the low training instances might cause under fitting of the learning algorithm.

We next focused on applying Deep Learning on the survival prediction. For this, we used a different approach where we can use the complete profile without performing any feature selection. We expected the DL model to learn the more important survival associated features by itself with this. Specially, gliomas are hard to be treated due to their heterogeneity. This heterogeneity occurs in the gene level. Heterogeneity means, from patient to another patient the genomic profiles vary. Thus, using the whole profile without selecting couple of features might enable the DL model to learn this heterogeneity more efficiently. Further, to identify the features that contributed for the model prediction, we used the SHAP explainability analysis. In addition, we explored the applicability of different loss functions to enhance the predictions.

For predicting the survival using DL, a superior performance was observed

with the normLSF in both prediction accuracy and model calibration. We can also summarize that LS enhances the classification network outcome with both CE loss and FL loss. Despite the finer performance, normLSF can add a little computational cost in model training. On the other hand, proposed cost function in equation 3.9 and 3.10 contains tunable parameters $\alpha, \gamma and \epsilon$ which may require additional adjustment. However, even following the best performed values of these parameters ($\alpha = 1, \gamma = 2.0 or 3.0 and \epsilon = 0.1$) in corresponding works such as FL [127] and LS [129] can improve the network performance.

Then, we develop a multi tasking model for survival and grade prediction of gliomas using the mutation profiles. To the best of my knowledge, this is the first work that have focused on multi tasking in prognosis prediction of gliomas. These two tasks, survival and the grade are closely associated factors of gliomas. Patients with higher grade gliomas, such as WHO grade IV Glioblastomas, have a lower survival, majority accounting for short survival category (<300 days). Therefore, application wise this multi task prediction creates more worth by enabling similar tasks to train parallely.

Finally the relationships between the genomic biomarkers, radiomic biomarkers and the survival of Glioblastoma patients were closely analysed. We demonstrate that the kurtosis feature extracted from the enhancement region has a postive correlation with the survival. Kurtosis is a dimension that depicts to which extend the intensity of that particular region is deviated from a perfectly normal Gaussian distribution. A similar association between the enhancement and the survival, was reported in previously done studies [154, 155].

Our next important observation is the expression level of COL3A1 gene's association with survival. It is a gene that had been previously identified as a prognostic marker of Glioblastomas and had shown a interconnection with survival [156]. We could also observe a correlation between the fractal dimension extracted from the necrosis region and the TIMP4 gene. Rorive et al [157] reports that the over expression of the TIMP4 gene can cause lower survival of Glioblastoma patients. Therefore, we can assume that we can predict the survival by extracting the fractal dimension during the initial screening of the glioma.

In addition, we could identify fractal dimension, which indicates the complexity of glioma subregion through boc counting approach, as a significanly different feature between the 4 subtypes of Glioblastoma.

Our method for subtype prediction with the radiomic features, provide a new direction for this glioma analysis research area. With the current drawbacks in the acquisition protocols, it is hard to achieve gene expression profiles with common background interventions. Therefore this method, which do not require invasive approaches to acquire biomarker profiles, can lead for more impactful prognosis algorithm development. The images are acquired at the early diagnosis stages, and features such as kurtosis, fractal dimension can be extracted after segmenting the tumor using automated DL based algorithms.

Verhaak et al [136] identify 4 Glioblastoma subtypes, which were grouped based on gene expression profiles. Howeverm the neural subtype was later recognized as it was determined through contamination with non tumor cells. Thus, Wang et al [158] found out Classical, Proneural and Mesenchymal as the three subtypes of glioblastomas. However, for our dataset we had only the Verhaak identified 4 subtype information. Therefore, we had to rely on those subtype information, which also had, now non existing Neural subtype. Therefore, we executed the experiments as a binary classification for each subtype separately.

In conclusion, our survival prediction accuracies, despite the data been used as the input, can be compared with the BraTS challenge, Survival prediction task performances. BraTS is a challenge that had been held for 10 consecutive in parallel with the MICCAI conference, in order to determine the finest AI based algorithm for glioma segmentation. In parallel, they have worked on determining the survival using the images. This can be considered as the benchmark for our survival prediction task.

To the best of our knowledge, this is the first work that have used genomics for survival prediction.

[1]https://www.cbica.upenn.edu/BraTS18/lboardValidationSurvival.html
[2]https://www.cbica.upenn.edu/BraTS19/lboardValidationSurvival.html
[3]https://www.cbica.upenn.edu/BraTS20/lboardValidationSurvival.html

Table 5.1: Comparison with existing best performing works.

| Work | Accuracy | MSE |
|---|---|---|
| Brat 2018 - 1st place [1] | 32.1% | 103839.359 |
| Brats 2019 - 1st place [2] | 58.6% | 105061.874 |
| Brats 2020 - 1st place [3] | 41.4% | 98704.655 |
| **Our work** | > 70% | - |

This year, in BraTS 2021, they have come up with a classification task for predicting the MGMT methylation status, using the MRI images. This implies the eager in the current exploration to identify the associations between images and the genomics. Thus, our work further stands out noteworthy among the state of art literature.

## 5.2 Limitations

In the present day, there are many researches getting published on glioma survival. However, reproducibility is questionable, in order to apply these research in clinical practise. In addition, to use survival analysis with radiomics in clinical practice, the radiomics should be able to reproduce and must be independent. According to the literature, these features can be influenced by many factors, including the imaging equipment, acquisition protocols, image preprocessing steps, and image reconstruction. These can affect survival prediction models or analysis tools. Hence, technical standardization has emerged into research studies, and clinical practice by associations such as Quantitative Imaging Biomarker Alliance, in order to ensure the intra and inter-machine reproducibility of radiomics. Moreover, the reproducibility of deep features, extracted from imaging data, is also a critical task and, thus Han et al. [41] incorporate deep features with handcrafted features to provide a higher accuracy for survival prediction. Nonetheless, genomics data adhere to the same limitation in reproducibility of curation and interpretation of genomic related analysis. Thus, organizations such as the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) have published the standards and guidelines for interpreting and reporting cancer related sequence variant, provid-

ing improvements for the field of precision medicine in cancer and genomic testing in practice. However, since we have used public data cohorts, we had to rely on their data collection and acquisition procedures followed, without focusing on the reproducibility from the scratch. This is a limitation of our study, which could have been overcome if we could use our own acquired data with a standard international acquisition protocol. Thus, this clinical applicability needs to focus on these protocols in the future.

More specifically our main focus of this research, the usability of genomics for survival prediction is a novel area, which does not have standards yet for gene expression profiling and mutation profiling from the biopsies/ glioma samples. This lead for inconsistencies between datasets, and ultimately lower testing performance, which is another limitation of our work.

We used publicly available, open-access datasets; The Cancer Genome Atlas (TCGA) and the Chinese Glioma Genome Atlas (CGGA). Nevertheless, these cohorts also consist of less than 500 glioma patient cases, due to the limitations in time and high cost in data acquisition. Therefore our experiments are more likely to underfit for the available data. Furthermore, training on just one dataset, can cause for overfitting for that particular dataset, making the models less sensitive for the unseen and new data.

Also, in the future, different genomic profiles from different platforms can be combined to create more accurate models, which is not currently possible due to the lack of data. In our datasets, the overlapping cases were not sufficient to train an AI algorithm.

## 5.3 Future Directions

This work is an initial level analysis on how we can develop tools for prognosis estimation in glioma patients. In order to make this clinical applicable, this field should be more closely analysed and must be highly accurate than this. Therefore, I believe future must mainly focus on improving the accuracy and the applicability by coming up with worldwide recognized acquisition protocols. Both

radiomics and genomics features must be more solidly extracted with automated algorithms, which can be done proposing standard feature extraction tools. This will definitely lead for more accurate survival prediction tool development in the future using AI algorithms.

In the present day, in clinical practise, tissue samples acquired using biopsy or surgery are used for genomic biomarker identification such as MGMT methylation, IDH1 mutation etc. In the future, this can be extended by obtaining high throughput profiles in clinical practise, thus, leading for these tools to be practically used.

Nonetheless, as the world is moving forward with the technology, especially startups like Panakeia [4] develop omics based AI platforms for diagnosis purposes. Thereby the future research must further focus on commercialization of these prognosis estimation tools.


[4] https://panakeia.ai

# Chapter 6

# CONCLUSION

Initially, our study has presented a solution for survival prediction of glioma patients based on genomics with a comparatively large dataset. We identified 7 gene signature associated with survival and proposed two approaches for prognosis prediction which have a potential ability to separate glioma patients into groups based on their survival and risk. We proposed a novel Bayesian neural network, for survival prediction that surpasses the state-of-art survival prediction approaches. Moreover, we established a comprehensive prognostic risk model based on Cox-PH, that estimates the risk of glioma patients including both GBM and LGG. Both of these approaches have promising predictive ability of survival and risk for glioma patients, with over 70% accuracy for overall survival class prediction. Next we focus on utilizing Deep Learning for survival prediction with gene expression profiles. We design a novel loss function, normLSF, by smoothing true labels and re-weighting the probability for easy and hard samples during the calculation of the loss. The proposed normLSF empirically showcases superior performance in terms of both prediction accuracy and confidence calibration with genomic data for glioma survival prediction. We use multiple genomic datasets of TCGA and CGGA, and the experimental results demonstrate that normLSF is improved the uncertainty calibration as well as prediction calibration. The models trained with our loss function are not only highly accurate but also determine the reliability of their prediction, which is very crucial for the computer-aided diagnosis system. Moreover, our loss function can alleviate the class-imbalance bias and enhances the accuracy explicitly for smaller datasets. To substantiate our proposition, we test our loss function on publicly available CIFAR100 dataset and observe the upgraded performances. By enabling well-calibration and uncertainty, we would like to focus on detecting out of distribution samples for out future work.

In addition, we address a solution for subtype prediction without employing

high dimensional genomics, that is also inconsistent with the platform acquired. We identified distinctive radiomics between subtypes, that reflects the underlying molecular level gene alterations between the subtypes. Accordingly, we developed a subtype predictive model, with machine learning for GBM patients. In conclusion, our work also delivers a promising subtype predictive model, that will lead for a better clinical maintenance of GBM patients.

In conclusion, this research explore various approaches for survival prediction with genomics. We believe this work can be further expanded by employing large datasets for more accurate algorithm development. In the future, this work will provide a guidance for clinically usable application development.

# References


[1] Quinn T Ostrom, Haley Gittleman, Gabrielle Truitt, Alexander Boscia, Carol Kruchko, and Jill S Barnholtz-Sloan. Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2011–2015. *Neuro-oncology*, 20(suppl_4):iv1–iv86, 2018.

[2] Quinn T Ostrom, Haley Gittleman, Jordonna Fulop, Max Liu, Rachel Blanda, Courtney Kromer, Yingli Wolinsky, Carol Kruchko, and Jill S Barnholtz-Sloan. Cbtrus statistical report: primary brain and central nervous system tumors diagnosed in the united states in 2008-2012. *Neuro-oncology*, 17(suppl_4):iv1–iv62, 2015.

[3] Quinn T Ostrom, Haley Gittleman, Peter Liao, Chaturia Rouse, Yanwen Chen, Jacqueline Dowling, Yingli Wolinsky, Carol Kruchko, and Jill Barnholtz-Sloan. Cbtrus statistical report: primary brain and central nervous system tumors diagnosed in the united states in 2007–2011. *Neuro-oncology*, 16(suppl_4):iv1–iv63, 2014.

[4] Michael Weller, Wolfgang Wick, Ken Aldape, Michael Brada, Mitchell Berger, Stefan M Pfister, Ryo Nishikawa, Mark Rosenthal, Patrick Y Wen, Roger Stupp, et al. Glioma. *Nature reviews Disease primers*, 1(1):1–18, 2015.

[5] Gregory N Fuller and Bernd W Scheithauer. The 2007 revised world health organization (who) classification of tumours of the central nervous system: newly codified entities. *Brain pathology*, 17(3):304–307, 2007.

[6] Felix Sahm, David Reuss, Christian Koelsche, David Capper, Jens Schittenhelm, Stephanie Heim, David TW Jones, Stefan M Pfister, Christel Herold-Mende, Wolfgang Wick, et al. Farewell to oligoastrocytoma: in situ molecular genetics favor classification as either oligodendroglioma or astrocytoma. *Acta neuropathologica*, 128(4):551–559, 2014.

[7] Kewal K Jain. A critical overview of targeted therapies for glioblastoma. *Frontiers in oncology*, 8:419, 2018.

[8] Hiroko Ohgaki and Paul Kleihues. Epidemiology and etiology of gliomas. *Acta neuropathologica*, 109(1):93–108, 2005.

[9] Pei Yang, Yongzhi Wang, Xiaoxia Peng, Gan You, Wei Zhang, Wei Yan, Zhaoshi Bao, Yinyan Wang, Xiaoguang Qiu, and Tao Jiang. Management and survival rates in patients with glioma in china (2004–2010): a retrospective study from a single-institution. *Journal of neuro-oncology*, 113(2):259–266, 2013.

[10] Elizabeth B Claus, Kyle M Walsh, John K Wiencke, Annette M Molinaro, Joseph L Wiemels, Joellen M Schildkraut, Melissa L Bondy, Mitchel Berger, Robert Jenkins, and Margaret Wrensch. Survival and low-grade glioma: the emergence of genetic information. *Neurosurgical focus*, 38(1):E6, 2015.

[11] Navodini Wijethilake, Dulani Meedeniya, Charith Chitraranjan, and Indika Perera. Survival prediction and risk estimation of glioma patients using mrna expressions. In *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 35–42. IEEE, 2020.

[12] Ping Wang, Yan Li, and Chandan K Reddy. Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, 51(6):1–36, 2019.

[13] L Ganau, M Paris, GK Ligarotti, and M Ganau. Management of gliomas: overview of the latest technological advancements and related behavioral drawbacks. *Behavioural neurology*, 2015, 2015.

[14] SE Metcalfe and R Grant. Biopsy versus resection for malignant glioma. *The Cochrane database of systematic reviews*, (3):CD002034–CD002034, 2001.

[15] N Upadhyay and AD3473894 Waldman. Conventional mri evaluation of gliomas. *The British journal of radiology*, 84(special_issue_2):S107–S111, 2011.

[16] P Kalavathi and VB Surya Prasath. Methods on skull stripping of mri head scan images—a review. *Journal of digital imaging*, 29(3):365–379, 2016.

[17] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341, 2012.

[18] Stefan Bauer, Thomas Fejes, and Mauricio Reyes. A skull-stripping filter for itk. *Insight Journal*, 2012, 2013.

[19] Jiangwei Lao, Yinsheng Chen, Zhi-Cheng Li, Qihua Li, Ji Zhang, Jing Liu, and Guangtao Zhai. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Scientific reports*, 7(1):1–8, 2017.

[20] N Wijethilake, M Islam, and H Ren. Radiogenomics model for overall survival prediction of glioblastoma. *Medical & Biological Engineering & Computing*, 58:1767–1777, 2020.

[21] Navodini Wijethilake, Mobarakol Islam, Dulani Meedeniya, Charith Chitraranjan, Indika Perera, and Hongliang Ren. Radiogenomics of glioblastoma: Identification of radiomics associated with molecular subtypes. In *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology*, pages 229–239. Springer, 2020.

[22] Florent Tixier, Hyemin Um, Dalton Bermudez, Aditi Iyer, Aditya Apte, Maya S Graham, Kathryn S Nevel, Joseph O Deasy, Robert J Young, and Harini Veeraraghavan. Preoperative mri-radiomics features improve prediction of survival in glioblastoma patients over mgmt methylation status alone. *Oncotarget*, 10(6):660, 2019.

[23] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D ac-

tive contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006.

[24] Yunguan Fu, Nina Montaña Brown, Shaheer U. Saeed, Adrià Casamitjana, Zachary M. C. Baum, Rémi Delaunay, Qianye Yang, Alexander Grimwood, Zhe Min, Stefano B. Blumberg, Juan Eugenio Iglesias, Dean C. Barratt, Ester Bonmati, Daniel C. Alexander, Matthew J. Clarkson, Tom Vercauteren, and Yipeng Hu. Deepreg: a deep learning toolkit for medical image registration. *Journal of Open Source Software*, 5(55):2705, 2020.

[25] Brian B Avants, Nick Tustison, and Gang Song. Advanced normalization tools (ants). *Insight j*, 2(365):1–35, 2009.

[26] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.

[27] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

[28] Maarouf A Hammoud, Raymond Sawaya, Weiming Shi, Peter F Thall, and Norman E Leeds. Prognostic significance of preoperative mri scans in glioblastoma multiforme. *Journal of neuro-oncology*, 27(1):65–73, 1996.

[29] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[31] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In *International MICCAI Brainlesion Workshop*, pages 287–297. Springer, 2017.

[32] Lucas Fidon, Sebastien Ourselin, and Tom Vercauteren. Generalized wasserstein dice score, distributionally robust deep learning, and ranger for brain tumor segmentation: Brats 2020 challenge. *arXiv preprint arXiv:2011.01614*, 2020.

[33] Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In *International MICCAI Brainlesion Workshop*, pages 311–320. Springer, 2018.

[34] Zeyu Jiang, Changxing Ding, Minfeng Liu, and Dacheng Tao. Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task. In *International MICCAI Brainlesion Workshop*, pages 231–241. Springer, 2019.

[35] Sangeetha Saman and Swathi Jamjala Narayanan. Survey on brain tumor segmentation and feature extraction of mr images. *International Journal of Multimedia Information Retrieval*, 8(2):79–99, 2019.

[36] Mobarakol Islam, V Jeya Maria Jose, and Hongliang Ren. Glioma Prognosis: Segmentation of the Tumor and Survival Prediction Using Shape, Geometric and Clinical Information. In Alessandro Crimi, Spyridon Bakas, Hugo Kuijf, Farahani Keyvan, Mauricio Reyes, and Theo van Walsum, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 142–153, Cham, 2019. Springer International Publishing.

[37] Parita Sanghani, Beng Ti Ang, Nicolas Kon Kam King, and Hongliang Ren. Overall survival prediction in glioblastoma multiforme patients from volumetric, shape and texture features using machine learning. *Surgical oncology*, 27(4):709–714, 2018.

[38] Xue Feng, Nicholas J Tustison, Sohil H Patel, and Craig H Meyer. Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features. *Frontiers in Computational Neuroscience*, 14:25, 2020.

[39] Li Sun, Songtao Zhang, Hang Chen, and Lin Luo. Brain tumor segmentation and survival prediction using multimodal mri scans with deep learning. *Frontiers in neuroscience*, 13:810, 2019.

[40] Joost JM Van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo JWL Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21):e104–e107, 2017.

[41] W Han, L Qin, C Bay, X Chen, K-H Yu, N Miskin, A Li, X Xu, and G Young. Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas. *American Journal of Neuroradiology*, 41(1):40–48, 2020.

[42] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[43] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.

[44] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *International MICCAI brainlesion workshop*, pages 178–190. Springer, 2017.

[45] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.

[46] Elodie Puybareau, Guillaume Tochon, Joseph Chazalon, and Jonathan Fabrizio. Segmentation of gliomas and prediction of patient overall survival: a

simple and fast procedure. In *International MICCAI Brainlesion Workshop*, pages 199–209. Springer, 2018.

[47] Lina Chato, Erik Chow, and Shahram Latifi. Wavelet transform to improve accuracy of a prediction model for overall survival time of brain tumor patients based on mri images. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 441–442. IEEE, 2018.

[48] Zeina A Shboul, Mahbubul Alam, Lasitha Vidyaratne, Linmin Pei, Mohamed I Elbakary, and Khan M Iftekharuddin. Feature-guided deep radiomics for glioblastoma patient survival prediction. *Frontiers in Neuroscience*, 13, 2019.

[49] Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *annual conference on medical image understanding and analysis*, pages 506–517. Springer, 2017.

[50] Dong Nie, Han Zhang, Ehsan Adeli, Luyan Liu, and Dinggang Shen. 3d deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In *International conference on medical image computing and computer-assisted intervention*, pages 212–220. Springer, 2016.

[51] Mobarakol Islam, V S Vibashan, V Jeya Maria Jose, Navodini Wijethilake, Uppal Utkarsh, and Hongliang Ren. Brain Tumor Segmentation and Survival Prediction Using 3D Attention UNet. In Alessandro Crimi and Spyridon Bakas, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 262–272, Cham, 2020. Springer International Publishing.

[52] Mariarosaria Incoronato, Marco Aiello, Teresa Infante, Carlo Cavaliere, Anna Maria Grimaldi, Peppino Mirabelli, Serena Monti, and Marco Salvatore. Radiogenomic analysis of oncological data: a technical survey. *International journal of molecular sciences*, 18(4):805, 2017.

[53] Alhasan Alkuhlani, Mohammad Nassef, and Ibrahim Farag. Multistage feature selection approach for high-dimensional cancer data. *Soft Computing*, 21(22):6895–6906, 2017.

[54] Christian Hartmann, Bettina Hentschel, Marcos Tatagiba, Johannes Schramm, Oliver Schnell, Clemens Seidel, Robert Stein, Guido Reifenberger, Torsten Pietsch, Andreas Von Deimling, et al. Molecular markers in low-grade gliomas: predictive or prognostic? *Clinical Cancer Research*, 17(13):4588–4599, 2011.

[55] Shuguang Zuo, Xinhong Zhang, and Liping Wang. A rna sequencing-based six-gene signature for survival prediction in patients with glioblastoma. *Scientific reports*, 9(1):1–10, 2019.

[56] Junmin Xian, Quanzhong Zhang, Xiwen Guo, Xiankun Liang, Xinhua Liu, and Yugong Feng. A prognostic signature based on three non-coding rna s for prediction of the overall survival of glioma patients. *FEBS open bio*, 9(4):682–692, 2019.

[57] Wen-Jing Zeng, Yong-Long Yang, Zheng-Zheng Liu, Zhi-Peng Wen, Yan-Hong Chen, Xiao-Lei Hu, Quan Cheng, Jian Xiao, Jie Zhao, and Xiao-Ping Chen. Integrative analysis of dna methylation and gene expression identify a three-gene signature for predicting prognosis in lower-grade gliomas. *Cellular Physiology and Biochemistry*, 47(1):428–439, 2018.

[58] Pooya Mobadersany, Safoora Yousefi, Mohamed Amgad, David A Gutman, Jill S Barnholtz-Sloan, José E Velázquez Vega, Daniel J Brat, and Lee AD Cooper. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences*, 115(13):E2970–E2979, 2018.

[59] Saima Rathore, Muhammad Aksam Iftikhar, and Zissimos Mourelatos. Prediction of overall survival and molecular markers in gliomas via anal-

ysis of digital pathology images using deep learning. *arXiv preprint arXiv:1909.09124*, 2019.

[60] Jingwei Zhao, Le Wang, Daliang Kong, Guozhang Hu, and Bo Wei. Construction of novel dna methylation-based prognostic model to predict survival in glioblastoma. *Journal of Computational Biology*, 2019.

[61] Wei-Zhen Gao, Lie-Mei Guo, Tian-Qi Xu, Yu-Hua Yin, and Feng Jia. Identification of a multidimensional transcriptome signature for survival prediction of postoperative glioblastoma multiforme patients. *Journal of translational medicine*, 16(1):368, 2018.

[62] Haley Gittleman, Andrew E Sloan, and Jill S Barnholtz-Sloan. An independently validated survival nomogram for lower-grade glioma. *Neuro-oncology*, 22(5):665–674, 2020.

[63] Niha Beig, Salendra Singh, Kaustav Bera, Prateek Prasanna, Gagandeep Singh, Jonathan Chen, Anas SaeedBamashmos, Addison Barnett, Kyle Hunter, Volodymyr Statsevych, et al. Sexually dimorphic radiogenomic models identify distinct imaging and biological pathways that are prognostic of overall survival in glioblastoma. *Neuro-oncology*, 2020.

[64] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

[65] L Ladha and T Deepa. Feature selection methods and algorithms. *International journal on computer science and engineering*, 3(5):1787–1797, 2011.

[66] Lei Yu and Huan Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pages 856–863, 2003.

[67] Jingwei Zhao, Le Wang, Daliang Kong, Guozhang Hu, and Bo Wei. Construction of novel dna methylation-based prognostic model to predict sur-

vival in glioblastoma. *Journal of Computational Biology*, 27(5):718–728, 2020.

[68] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.

[69] Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

[70] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

[71] Kyrre E Emblem, Marco C Pinho, Frank G Zöllner, Paulina Due-Tonnessen, John K Hald, Lothar R Schad, Torstein R Meling, Otto Rapalino, and Atle Bjornerud. A generic support vector machine model for preoperative glioma survival associations. *Radiology*, 275(1):228–234, 2015.

[72] Luke Macyszyn, Hamed Akbari, Jared M Pisapia, Xiao Da, Mark Attiah, Vadim Pigrish, Yingtao Bi, Sharmistha Pal, Ramana V Davuluri, Laura Roccograndi, et al. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. *Neuro-oncology*, 18(3):417–425, 2015.

[73] Faisal M Khan and Valentina Bayer Zubek. Support vector regression for censored data (svrc): a novel tool for survival analysis. In *2008 Eighth IEEE International Conference on Data Mining*, pages 863–868. IEEE, 2008.

[74] E Marubini, A Morabito, and MG Valsecchi. Prognostic factors and risk groups: some results given by using an algorithm suitable for censored survival data. *Statistics in medicine*, 2(2):295–303, 1983.

[75] A Ciampi, RS Bush, M Gospodarowicz, and JE Till. An approach to classifying prognostic factors related to survival experience for non-hodgkin's

lymphoma patients: Based on a series of 982 patients: 1967–1975. *Cancer*, 47(3):621–627, 1981.

[76] Louis Gordon and Richard A Olshen. Tree-structured survival analysis. *Cancer treatment reports*, 69(10):1065–1069, 1985.

[77] Antonio Ciampi, Johanne Thiffault, Jean-Pierre Nakache, and Bernard Asselain. Stratification by stepwise regression, correspondence analysis and recursive partition: a comparison of three methods of analysis for survival data with covariates. *Computational statistics & data analysis*, 4(3):185–204, 1986.

[78] A Ciampi. Recursive partition and amalgamation (recpam) for censored survival data: criteria for tree selection. *Statistical Software Newsletter*, 14(2):78–81, 1988.

[79] Walter J Curran Jr, Charles B Scott, John Horton, James S Nelson, Alan S Weinstein, A Jennifer Fischbach, Chu H Chang, Marvin Rotman, Sucha O Asbell, Robert E Krisch, et al. Recursive partitioning analysis of prognostic factors in three radiation therapy oncology group malignant glioma trials. *JNCI: Journal of the National Cancer Institute*, 85(9):704–710, 1993.

[80] Glenn Bauman, Knut Lote, David Larson, Lukas Stalpers, Christopher Leighton, Barbara Fisher, William Wara, David Macdonald, Larry Stitt, and J Gregory Cairncross. Pretreatment factors predict overall survival for patients with low-grade glioma: a recursive partitioning analysis. *International Journal of Radiation Oncology* Biology* Physics*, 45(4):923–929, 1999.

[81] H. J. Cho and S. Hong. Median regression tree for analysis of censored survival data. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 38(3):715–726, 2008.

[82] María L Gandía-González, Sebastián Cerdán, Laura Barrios, Pilar López-Larrubia, Pablo G Feijoó, Alexis Palpan Jr, José M Roda, and Juan So-

livera. Assessment of overall survival in glioma patients as predicted by metabolomic criteria. *Frontiers in Oncology*, 9:328, 2019.

[83] J Ross Quinlan et al. Bagging, boosting, and c4. 5. In *AAAI/IAAI, Vol. 1*, pages 725–730, 1996.

[84] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[85] Bradley Efron. Tibshirani,(1993), an introduction to the bootstrap. *Monographs on Statistics and Applied Probability. Chapman & Hall/CRC*, 1, 1960.

[86] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.

[87] Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, Michael S Lauer, et al. Random survival forests. *The annals of applied statistics*, 2(3):841–860, 2008.

[88] Yoon Seong Choi, Sung Soo Ahn, Jong Hee Chang, Seok-Gu Kang, Eui Hyun Kim, Se Hoon Kim, Rajan Jain, and Seung-Koo Lee. Machine learning and radiomic phenotyping of lower grade gliomas: improving survival prediction. *European Radiology*, pages 1–9, 2020.

[89] Yoav Freund and Robert E Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory*, pages 23–37. Springer, 1995.

[90] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[91] Torsten Hothorn, Peter Bühlmann, Sandrine Dudoit, Annette Molinaro, and Mark J Van Der Laan. Survival ensembles. *Biostatistics*, 7(3):355–373, 2006.

[92] Jun Zhu, Jianfei Chen, Wenbo Hu, and Bo Zhang. Big learning with bayesian methods. *National Science Review*, 4(4):627–651, 2017.

[93] Simon Rabinowicz, Arjen Hommersom, Raphaela Butz, and Matt Williams. A prognostic model of glioblastoma multiforme using survival bayesian networks. In *Conference on Artificial Intelligence in Medicine in Europe*, pages 81–85. Springer, 2017.

[94] Mu Zhou, Baishali Chaudhury, Lawrence O Hall, Dmitry B Goldgof, Robert J Gillies, and Robert A Gatenby. Identifying spatial imaging biomarkers of glioblastoma multiforme for survival group prediction. *Journal of Magnetic Resonance Imaging*, 46(1):115–123, 2017.

[95] Vinicius Bonato, Veerabhadran Baladandayuthapani, Bradley M Broom, Erik P Sulman, Kenneth D Aldape, and Kim-Anh Do. Bayesian ensemble methods for survival prediction in gene expression data. *Bioinformatics*, 27(3):359–367, 2011.

[96] Stephen P Luttrell. A bayesian analysis of self-organizing maps. *Neural Computation*, 6(5):767–794, 1994.

[97] Pedro Henriques Abreu, Miriam Seoane Santos, Miguel Henriques Abreu, Bruno Andrade, and Daniel Castro Silva. Predicting breast cancer recurrence using machine learning techniques: a systematic review. *ACM Computing Surveys (CSUR)*, 49(3):1–40, 2016.

[98] Sun-Mi Lee and Patricia A Abbott. Bayesian networks for knowledge discovery in large datasets: basics for nurse researchers. *Journal of biomedical informatics*, 36(4-5):389–399, 2003.

[99] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.

[100] Mu Zhou, Lawrence O Hall, Dmitry B Goldgof, Robert J Gillies, and Robert A Gatenby. Survival time prediction of patients with glioblastoma multiforme tumors using spatial distance measurement. In *Medical Imaging 2013: Computer-Aided Diagnosis*, volume 8670, page 86702O. International Society for Optics and Photonics, 2013.

[101] Stephen R Piccolo and Lewis J Frey. Clinical and molecular models of glioblastoma multiforme survival. *International journal of data mining and bioinformatics*, 7(3):245, 2013.

[102] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[103] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[104] David Faraggi and Richard Simon. A neural network model for survival data. *Statistics in medicine*, 14(1):73–82, 1995.

[105] Travers Ching, Xun Zhu, and Lana X Garmire. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS computational biology*, 14(4):e1006076, 2018.

[106] Safoora Yousefi, Fatemeh Amrollahi, Mohamed Amgad, Chengliang Dong, Joshua E Lewis, Congzheng Song, David A Gutman, Sameer H Halani, Jose Enrique Velazquez Vega, Daniel J Brat, et al. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. *Scientific reports*, 7(1):1–11, 2017.

[107] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[108] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[109] Justin Bo-Kai Hsu, Tzu-Hao Chang, Gilbert Aaron Lee, Tzong-Yi Lee, and Cheng-Yu Chen. Identification of potential biomarkers related to glioma survival by gene expression profile analysis. *BMC medical genomics*, 11(7):34, 2019.

[110] Lingchen Wang, Zhengwei Yan, Xiaona He, Cheng Zhang, Huiqiang Yu, and Quqin Lu. A 5-gene prognostic nomogram predicting survival probability of glioblastoma patients. *Brain and behavior*, 9(4):e01258, 2019.

[111] Tong Yang, Ping Mao, Xianhai Chen, Xuan Niu, Gaofeng Xu, Xiaobin Bai, and Wanfu Xie. Inflammatory biomarkers in prognostic analysis for patients with glioma and the establishment of a nomogram. *Oncology letters*, 17(2):2516–2522, 2019.

[112] Farnoosh Abbas-Aghababazadeh, Qian Li, and Brooke L Fridley. Comparison of normalization approaches for gene expression studies completed with high-throughput sequencing. *PloS one*, 13(10), 2018.

[113] Amalia Annest, Roger E Bumgarner, Adrian E Raftery, and Ka Yee Yeung. Iterative bayesian model averaging: A method for the application of survival analysis to high-dimensional microarray data. *BMC bioinformatics*, 10(1):72, 2009.

[114] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[115] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. In *Advances in neural information processing systems*, pages 6638–6648, 2018.

[116] Iresha Rubasinghe and Dulani Meedeniya. Ultrasound nerve segmentation using deep probabilistic programming. *Journal of ICT Research and Applications*, 13(3):241–256, 2019.

[117] Eli Bingham, Jonathan P Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D Goodman. Pyro: Deep universal probabilistic pro-

gramming. *The Journal of Machine Learning Research*, 20(1):973–978, 2019.

[118] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[119] Terry M Therneau. *A Package for Survival Analysis in R*, 2020. R package version 3.2-3.

[120] Zheng Zhao, Kenan Zhang, Qiangwei Wang, Guanzhang Li, Fan Zeng, Ying Zhang, Fan Wu, Ruichao Chai, Zheng Wang, Chuanbao Zhang, et al. Chinese glioma genome atlas (cgga): A comprehensive resource with functional genomic data for chinese glioma patients. *bioRxiv*, 2020.

[121] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[122] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[123] Mingxing Tan and Quoc V Le. Mixconv: Mixed depthwise convolutional kernels. *arXiv*, pages arXiv–1907, 2019.

[124] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Zhi Zhang, Haibin Lin, Yue Sun, Tong He, Jonas Mueller, R Manmatha, et al. Resnest: Split-attention networks. *arXiv preprint arXiv:2004.08955*, 2020.

[125] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Pro-*

*ceedings of the IEEE conference on computer vision and pattern recognition*, pages 6848–6856, 2018.

[126] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[127] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[128] Bjorn Barz and Joachim Denzler. Deep learning on small datasets without pre-training using cosine loss. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 1371–1380, 2020.

[129] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[130] Rafael Müller, Simon Kornblith, and Geoffrey E Hinton. When does label smoothing help? In *Advances in Neural Information Processing Systems*, pages 4694–4703, 2019.

[131] Yu Zhang and Qiang Yang. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[132] Yan Li, Jie Wang, Jieping Ye, and Chandan K Reddy. A multi-task learning formulation for survival analysis. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1715–1724, 2016.

[133] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

[134] Mobarakol Islam, VS Vibashan, and Hongliang Ren. Ap-mtl: Attention pruned multi-task learning model for real-time instrument detection and segmentation in robot-assisted surgery. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8433–8439. IEEE, 2020.

[135] Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *International Conference on Machine Learning*, pages 794–803. PMLR, 2018.

[136] Roel GW Verhaak, Katherine A Hoadley, Elizabeth Purdom, Victoria Wang, Yuan Qi, Matthew D Wilkerson, C Ryan Miller, Li Ding, Todd Golub, Jill P Mesirov, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in pdgfra, idh1, egfr, and nf1. *Cancer cell*, 17(1):98–110, 2010.

[137] Steve Pieper, Michael Halle, and Ron Kikinis. 3d slicer. In *2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No. 04EX821)*, pages 632–635. IEEE, 2004.

[138] Abhijit Guha Roy, Nassir Navab, and Christian Wachinger. Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 421–429. Springer, 2018.

[139] Kourosh M Naeini, Whitney B Pope, Timothy F Cloughesy, Robert J Harris, Albert Lai, Ascia Eskin, Reshmi Chowdhury, Heidi S Phillips, Phioanh L Nghiemphu, Yalda Behbahanian, et al. Identifying the mesenchymal molecular subtype of glioblastoma using quantitative volumetric analysis of anatomic magnetic resonance images. *Neuro-oncology*, 15(5):626–634, 2013.

[140] Laura Follia, Giulio Ferrero, Giorgia Mandili, Marco Beccuti, Daniele Giordano, Rosella Spadi, Maria Antonietta Satolli, Andrea Evangelista, Hi-

royuki Katayama, Wang Hong, et al. Integrative analysis of novel metabolic subtypes in pancreatic cancer fosters new prognostic biomarkers. *Frontiers in oncology*, 9:115, 2019.

[141] Evgenia Dimitriadou, Kurt Hornik, Friedrich Leisch, David Meyer, Andreas Maintainer, friedrich Leisch@ci, Ac Tuwien, and At. The e1071 package. 08 2006.

[142] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1):10–18, 2009.

[143] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. *arXiv preprint arXiv:1706.04599*, 2017.

[144] Mahdi Pakdaman Naeini, Gregory Cooper, and Milos Hauskrecht. Obtaining well calibrated probabilities using bayesian binning. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[145] Jeremy Nixon, Michael W Dusenberry, Linchuan Zhang, Ghassen Jerfel, and Dustin Tran. Measuring calibration in deep learning. In *CVPR Workshops*, pages 38–41, 2019.

[146] Arsenii Ashukha, Alexander Lyzhov, Dmitry Molchanov, and Dmitry Vetrov. Pitfalls of in-domain uncertainty estimation and ensembling in deep learning. *arXiv preprint arXiv:2002.06470*, 2020.

[147] Max-Heinrich Laves, Sontje Ihler, Karl-Philipp Kortmann, and Tobias Ortmaier. Well-calibrated model uncertainty with temperature scaling for dropout variational inference. *arXiv preprint arXiv:1909.13550*, 2019.

[148] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in neural information processing systems*, pages 4765–4774, 2017.

[149] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825*, 2017.

[150] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. *arXiv preprint arXiv:1703.01365*, 2017.

[151] Jie Ren, Meiqing Lou, Jinlong Shi, Yajun Xue, and Daming Cui. Identifying the genes regulated by idh1 via gene-chip in glioma cell u87. *International journal of clinical and experimental medicine*, 8(10):18090, 2015.

[152] Shanrong Zhao, Wai-Ping Fung-Leung, Anton Bittner, Karen Ngo, and Xuejun Liu. Comparison of rna-seq and microarray in transcriptome profiling of activated t cells. *PloS one*, 9(1):e78644, 2014.

[153] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):594–611, 2006.

[154] JA1 Carrillo, A Lai, PL Nghiemphu, HJ Kim, HS Phillips, S Kharbanda, P Moftakhar, S Lalaezari, W Yong, BM Ellingson, et al. Relationship between tumor enhancement, edema, idh1 mutational status, mgmt promoter methylation, and survival in glioblastoma. *American Journal of Neuroradiology*, 33(7):1349–1355, 2012.

[155] David A Gutman, Lee AD Cooper, Scott N Hwang, Chad A Holder, JingJing Gao, Tarun D Aurora, William D Dunn Jr, Lisa Scarpace, Tom Mikkelsen, Rajan Jain, et al. Mr imaging predictors of molecular profile and survival: multi-institutional study of the tcga glioblastoma data set. *Radiology*, 267(2):560–569, 2013.

[156] Yuan-Feng Gao, Xiao-Yuan Mao, Tao Zhu, Chen-Xue Mao, Zhi-Xiong Liu, Zhi-Bin Wang, Ling Li, Xi Li, Ji-Ye Yin, Wei Zhang, et al. Col3a1 and snap91: novel glioblastoma markers with diagnostic and prognostic value. *Oncotarget*, 7(43):70494, 2016.

[157] Sandrine Rorive, Xavier Moles Lopez, Calliope Maris, Anne-Laure Trepant, Sébastien Sauvage, Niloufar Sadeghi, Isabelle Roland, Christine Decaestecker, and Isabelle Salmon. Timp-4 and cd63: new prognostic biomarkers in human astrocytomas. *Modern pathology*, 23(10):1418–1428, 2010.

[158] Qianghu Wang, Baoli Hu, Xin Hu, Hoon Kim, Massimo Squatrito, Lisa Scarpace, Ana C deCarvalho, Sali Lyu, Pengping Li, Yan Li, et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer cell*, 32(1):42–56, 2017.