IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 23, 2021, accepted March 10, 2021, date of publication March 15, 2021, date of current version March 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3065965

Glioma Survival Analysis Empowered With Data Engineering–A Survey

NAVODINI WIJETHILAKE[1], (Student Member, IEEE), DULANI MEEDENIYA[1], (Member, IEEE), CHARITH CHITRARANJAN[1], INDIKA PERERA[1], (Senior Member, IEEE), MOBARAKOL ISLAM[2], (Member, IEEE), AND HONGLIANG REN[3,4], (Senior Member, IEEE)

[1]Department of Computer Science and Engineering, University of Moratuwa, Moratuwa 10400, Sri Lanka
[2]Biomedical Image Analysis Group, Imperial College London, London SW7 2AZ, U.K.
[3]Department of Biomedical Engineering, National University of Singapore, Singapore 117583
[4]Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong

Corresponding author: Dulani Meedeniya (dulanim@cse.mrt.ac.lk)

This work was supported by the University of Moratuwa, Sri Lanka under the Conference & Publishing Grant.

ABSTRACT Survival analysis is a critical task in glioma patient management due to the inter and intra tumor heterogeneity. In clinical practice, clinicians estimate the survival with their experience, which can be biased and optimistic. Over the past decades, diverse survival analysis approaches were proposed incorporating distinct data such as imaging and genetic information. The remarkable advancements in imaging and high throughput omics and sequencing technologies have enabled the acquisition of this information of glioma patients efficiently, providing novel insights for survival estimation in the present day. Besides, in the past years, machine learning techniques and deep learning have emerged into the field of survival analysis of glioma patients trading off the traditional statistical analysis-based survival analysis approaches. In this survey paper, we explore the prognostic parameters acquired, utilizing diagnostic imaging techniques and genomic platforms for survival or risk estimation of glioma patients. Further, we review the techniques, learning and statistical analysis algorithms, along with their benefits and limitations used for prognosis prediction. Consequently, we highlight the challenges of the existing state-of-the-art survival prediction studies and propose future directions in the field of research.

INDEX TERMS Survival prediction, risk analysis, glioma, genomics, radiomics, radiogenomics, prognosis.

I. INTRODUCTION

A. GENERAL OVERVIEW OF GLIOMAS AND SUBTYPES

Gliomas are tumors that occur in glial or other progenitor cells. Gliomas account for 26.7% of all primary brain and central nervous system tumors. Generally, gliomas occur within the brain, mostly in the frontal, parietal, temporal lobes, and rarely in the occipital lobe. They also develop in spinal code and cauda equina cerebrum [1]. Based on the histology, Glioblastomas (GBM) account for 54.7% of primary brain and central nervous systems gliomas. Astrocytomas and GBMs combined account for 75% of gliomas. The incidence rate of gliomas decreases with the increasing age for the children and adolescents (age between 15-19 years), and approximately 46.5% of tumors are gliomas in this particular age group [2]. However, the incidence rate increases for the patients over 20 years and the highest is among the age of 85+ years. The median age at the diagnosis of GBM is 65 years. Further, GBM, astrocytomas, and oligodendrogliomas are comparatively higher in males with light skin colours than females [3].

The initial glioma classifications consider the underlying histology based morphological appearances in particular cell types. The brain cells have intensive networks that help to maintain the functions of the human brain [4].

- Astrocytes: These are the main connective tissue cells that can be found in the brain. When these cells show morphological similarities in a specific region, they are called astrocytomas.
- Oligodendrocytes: These cells wrap the neuronal axons in the brain with a myelin sheath. Oligodendrogliomas derive from these cells.
- Ependymal cells: Ependydoma occurs in these cells, which occurs less frequently in humans.

The associate editor coordinating the review of this manuscript and approving it for publication was Hao Luo.

World Health Organization (WHO) has categorized the astrocytomas, oligodendrogliomas, oligoastrocytomas, and ependymal gliomas into four grades based on the natural disease cause, absence, and presence of anaplastic features and malignancy [5].

- WHO grade I: Least malignant behavior and slow-growing. Pilocytic Astrocytomas belong to this.
- WHO grade II: Slow growing, but sometimes, a brain-invasive growth can be seen. Diffuse astrocytoma belongs to this category.
- WHO grade III: Rapidly growing gliomas, with histological features of anaplasia.
- WHO grade IV: Most malignant glioma, known as GBM and distinguished by the presence of necrosis and microvascular proliferation.

Recently, the underlying molecular pathogenesis is analyzed to identify genetic alterations, which can cause gliomas. This is slightly complementary to histological classifications and diagnostics.

The combination of WHO grade II and III gliomas are often referred to as Lower Grade Gliomas (LGG). Although the traditional classification of gliomas based on histology separates glioma into astrocytomas, oligodendrogliomas, and mixed oligoastrocytomas, these three can be categorized into 2 subtypes based on their molecular profiles.

- Mutation in tumor protein p53 (TP53) gene along with the Mutation in α-thalassemia/mental retardation syndrome X-linked (ATRX) gene: These gene markers indicate 'astrocytic' genotype.
- Mutation in telomerase reverse transcriptase (TERT) promoter and co-deletion of chromosomal arms 1p and 19q: These gene markers indicate the 'oligodendroglial' genotype.

Therefore, the existence of oligoastrocytomas histological type is not required any more [6]. Both of these genotypes frequently show mutations in Isocitrate dehydrogenase 1 (IDH1) and sometimes in Isocitrate dehydrogenase 2 (IDH2). Further, the novel classification of gliomas is into three groups based on the outcomes, natural histories, and the response for the treatments [4].

- IDH1 mutant, 1p/19q co-deleted tumors are mostly oligodendroglial histology: these tumors have the best prognosis/ longer survival.
- IDH1 mutant, 1p/19q non co-deleted tumors that are mostly astrocytic morphology: these tumors have an intermediate survival
- IDH wild-type, 1p/19q non co-deleted tumors: these have a poor prognosis/ short survival.

Gliomas, based on their histology, localization, and growth, have different levels of mortality and morbidity. GBM, the most malignant glioma, has only a median survival of 15 months, even with surgical resection followed by radiation therapy and chemotherapy [7]. In fact, less than 3% of GBM patients survived after 5 years from the initial diagnosis [8]. Moreover, WHO grade II and WHO grade III gliomas have median overall survival times of 78.1 and 37.6 months, respectively [9]. Considering the histology, the median survival of astrocytomas and oligodendrogliomas are 5.2 and 7.2 years, respectively [10].

B. IMPORTANCE OF SURVIVAL ESTIMATION IN GLIOMA

Overall survival of glioma patients is measured either from the date of diagnosis of glioma or from the date the treatments begin. The length of the time is measured until the patient is alive. This measurement is frequently used to decide the impact of a particular treatment given to the patient. The clinical protocols are planned to maximize the overall survival always without compromising the quality of life of patients. The 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, which ultimately impact the survival of patients [4]. These factors 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 [11]. Moreover, standard treatments such as anti-seizure agents and corticosteroids, used in routine medications, cause nausea, fatigue, drowsiness, and many more short term side effects. Meanwhile, they can induce long term effects such as diabetes mellitus, depression, hypertension, and osteoporosis [4].

Moreover, other than physical consequences, non-physical impairments arise in spiritual, psychological, behavioral domains. Nevertheless, the families 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 long term survival with minimum side effects. Sometimes, the patients will weaken due to side effects than the effects of the glioma. More often, clinicians estimate the survival of patients by clinical factors assessment, imaging assessment, and their experience. Researches show that these decisions can be biased, inaccurate, and optimistic [12]. This will also impact the patient and their families to face the situation and prepare for future challenges.

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 [13].

C. CURRENT ISSUES

Traditionally, the research on survival prediction of gliomas with medical images and computational modelling is mostly performed with the acquisition of images, such as Magnetic Resonance Images (MRI), Computed Tomography (CT), or Positron Emission Tomography (PET), followed by preprocessing, feature extraction, and classification into short, medium and long survival groups [14], [15]. Initially,

semi-automated approaches were popular, where segmentation and shape, histogram, and volume feature extraction were performed manually, and machine learning (ML) based classifications were performed for survival estimation. State-of-the-art, fully automated deep learning (DL) based segmentation followed by feature extraction and ML based classification and regression models are the most common approaches with a maximum accuracy reported around 70% [16]. These imaging modalities are still unable to capture the intra-tumor heterogeneity, which occurs at the molecular level; although, imaging approaches have enlightened the survival prediction research through its non-invasive behaviour.

Mostly in gliomas, statistical analysis and risk estimation approaches are proposed to estimate the survival and risk. The risk is given as a risk score that involves expression values of few genes, or any other factors related to survival and even can be measured using nomograms. However, these approaches are also unable to accurately reflect the survival of glioma patients. Also, there are various methodologies proposed by different research teams, that are independent and unique, making it more complicated to compare their approaches and apply in the clinical domain.

D. MOTIVATION

The new improvements in the high throughput technologies such as microarray and DNA/RNA sequencing have made the acquisition of heterogeneous genomic features credible and thus, more accurate survival analysis can be conducted, overcoming the drawbacks in the traditional radiomic approaches [17]. Nevertheless, the advancements in fields such as Artificial Intelligence, with the capability of handling high dimensional data and models, has arisen in the field of patient management related to cancer [18]. This can provide an absolute orientation for the future research of glioma prognosis estimation. Therefore, to incorporate the novel approaches, it is necessary to be aware of the current state-of-the-art techniques while identifying their ambiguities and inadequacies.

In addition, the current exploration is not yet feasible to be implemented in glioma clinical management. Through this work, we were motivated to explore the current approaches for survival prediction of gliomas, focusing on radiomics, genomics and other survival related data. Consequently, we identify the limitations and challenges that might gravitate the state-of-the-art survival prediction approaches for a clinically plausible era. Thus, we believe this survey will motivate many researchers to find directions to produce more precise patient survival estimation models.

E. CONTRIBUTION

This study mainly explores the current survival analysis techniques such as ML algorithms, DL approaches and statistical analysis techniques. Our first objective is to identify the different types of data, including imaging features and genomic data, used in the survival prediction of gliomas. Moreover, we compare the preprocessing and other techniques by analyzing the accuracy and limitations in the existing studies. As another objective, we analyze the ML and DL algorithms that have been used for survival prediction with distinct data. Next, we review the statistical analysis based methods, apart from learning techniques, that are being used for survival or risk assessment of glioma patients. Ultimately, we report the limitations of the current research that can be addressed in the future studies of glioma survival prediction, providing an explicit perception of clinical execution.We suppose that this research will broaden the horizons of the field of survival prediction of glioma patients forging ahead to an effectual glioma management.

This study is important for researchers who are engaged in developing tools and algorithms for survival estimation of glioma and various other cancer types to get a general idea about the state-of-the-art methods. Other than that, this is important for clinicians to decide the capability of utilizing the available algorithms in clinical practice, and to provide the directions required for the development in this field, further adding clinical value and importance.

II. DATA TYPES USED IN GLIOMA SURVIVAL ANALYSIS

A. GLIOMA SCREENING AND DATA COLLECTION IN CLINICAL PRACTICE

The screening of gliomas is frequently done using several types of tests. Physical examining is the initial test done by clinicians, where the medical history, family history and lifestyles are assessed closely. As neurological tests, a set of questions are asked from the patients to determine the changes that have occurred in the brain, spinal cord and nerves with the progression of the tumor [19]. Visual field tests are also commonly performed in the initial diagnosis process, since eyesight can get disrupted with the muscles being compressed due to the tumor within the brain. After these types of initial screening normally imaging tests are obtained. The most frequently used imaging tests are Computed Tomography Scan (CT Scans), Magnetic Resonance Imaging (MRI), Single Photon Emission Tomography Scan (SPECT), Positron Emission Tomography (PET) and Angiograms. Other than that, blood and urine tests are done to identify the substances, that change as an effect of tumor [19].

However, the cancerous status of the tumor has to be determined by examining the cells using a microscope. For this, biopsy or surgical resection are performed. In biopsy, only a piece of tissue from the tumor region is acquired and in surgical resection, the complete tumor is removed with the help of images acquired from the tumor region [20]. With the tumor tissue sample received with one of the above methods, the pathologists closely observe the cells, i.e. the pathological images, through Light & Electron Microscopes, to verify the signs of cancer and further, genomic profiles are acquired. These genomic profiles include the gene expression, methylation and mutation profiles, which we further discuss in Section II-C.

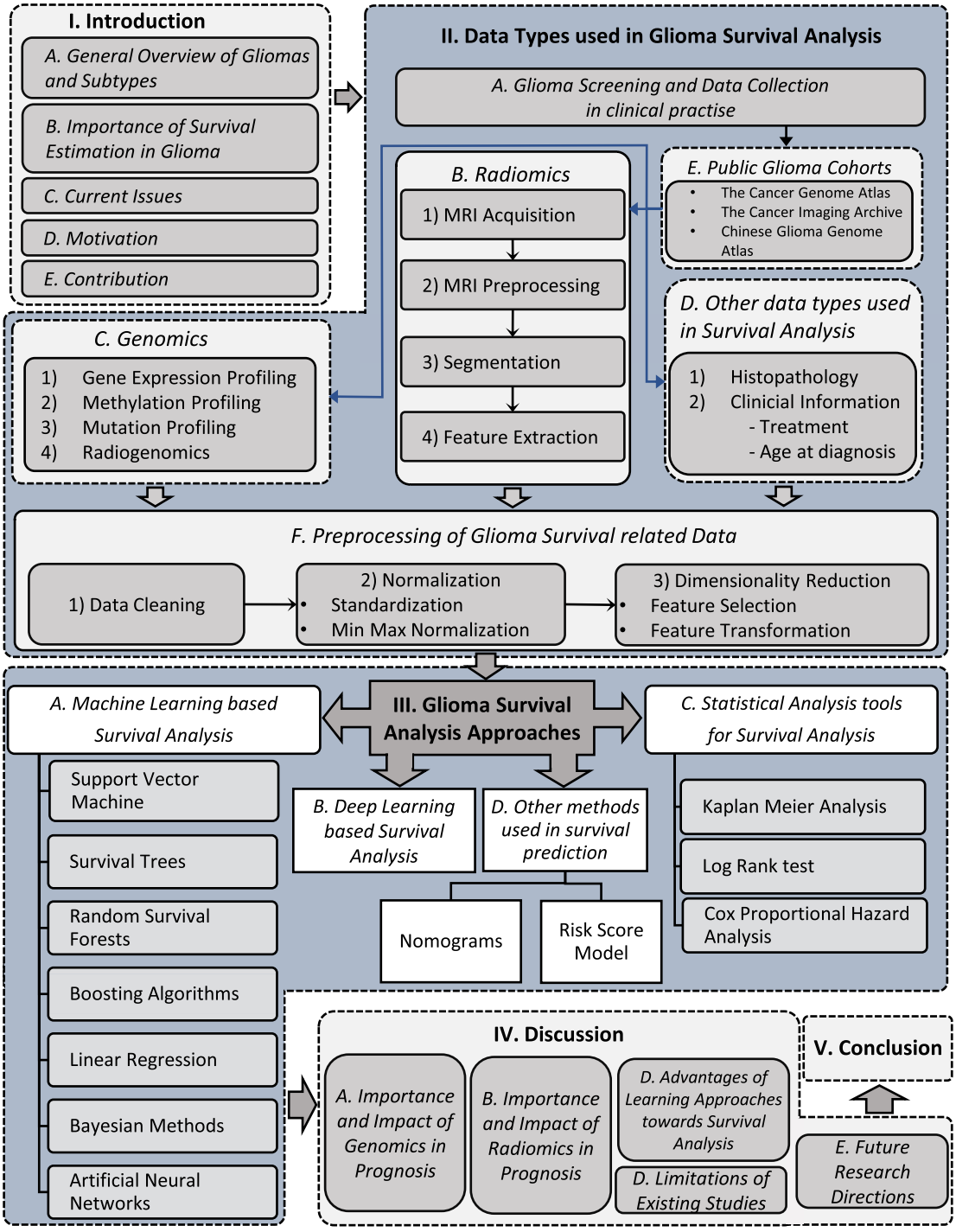

FIGURE 1. Overview of this survey paper on glioma prognosis.

Consequently, all these types of test profiles are achieved from glioma patients, throughout the diagnosis and screening processes. Thus, many tests are collected from each patient before diagnosis and after the surgery, and these records are collected by many institutions for research and various other purposes. These records are used to develop various prognostic tools and algorithms, to provide precise personalized disease management as we discuss in this work. Some institutions have made these details available to the public with open access as explained in Section II-E.

B. RADIOMICS

As discussed in Section II-A, different imaging techniques are used for clinical purposes in practice. However, for

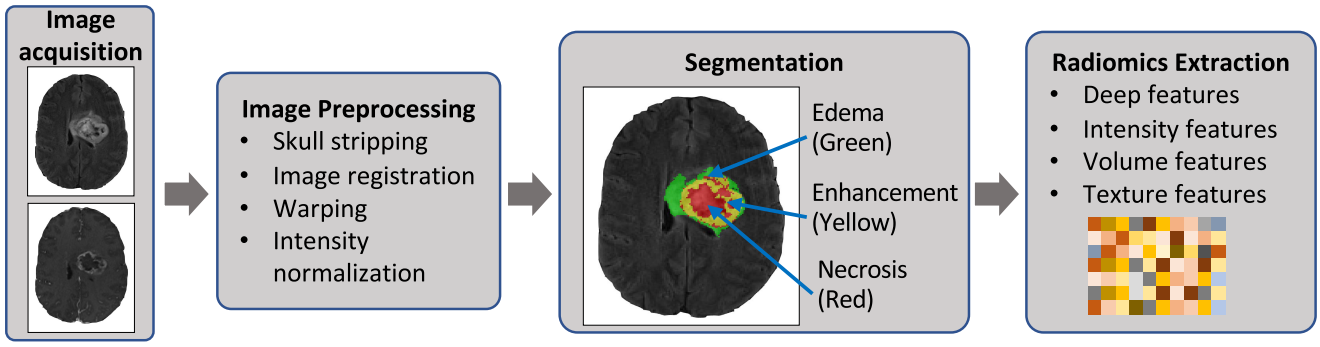

FIGURE 2. A general radiomics extraction pipeline.

TABLE 1. Open-source Tools available for Pre-processing and Segmentation of MRI.

Tool	Available Features	Link
ITK-SNAP [29]	Manual Segmentation	http://www.itksnap.org
3D slicer [23]	Skull stripping, Image registration, Intensity normalization, Segmentation, Visualization	https://www.slicer.org
DeepReg [30]	DL based Registration	https://github.com/DeepRegNet/DeepReg
ANTs [31]	Image Segmentation, Image Registration	https://github.com/ANTsX/ANTs
FSL [32]	MRI Segmentation, Registration	https://fsl.fmrib.ox.ac.uk/
FreeSurfer [33]	Skull Stripping, MRI Registration, Segmentation, Reconstruction	https://surfer.nmr.mgh.harvard.edu

survival related analysis these images have to be processed to extract imaging features from the tumor region. In this section, we discuss the process of acquiring those features step by step. A general pipeline for extracting radiomics from MRI sequences is shown in Figure 2. This has 4 main steps; 1) image acquisition: MRI scans are acquired, 2) image preprocessing: the images are skull stripped, registered and normalized, 3) image segmentation: the tumor region is segmented into subregions, and finally, 4) Radiomic feature extraction: various image related features are extracted. These four steps are explained as follows.

1) MRI ACQUISITION

The conventional state-of-the-art approach for survival prediction is the radiology imaging-based method. As a non-invasive tool in cancer clinical protocol, structural MRI is widely used to capture morphological, functional, and structural information related to cancer. Mostly, T1-weighted (T1W) and T2-weighted (T2W) that shows the basic pulse sequences in MRI, Fluid-attenuated inversion recovery (FLAIR), and gadolinium-enhanced sequences of MRI are used for diagnosis, surveillance, and monitoring of gliomas. Specifically, the pathological enhancement, which is common is malignant gliomas, can be denoted by acquiring and assessing FLAIR and contrast-enhanced T1 weighted imaging sequences [21]. There are several publicly available datasets with MRI sequences and corresponding clinical information for survival analysis, which we discuss in Section II-E. Other than that, many institutions use their own private MRI cohorts, collected and processed by themselves for prognosis studies.

2) MRI PREPROCESSING

A major obstacle for performing segmentation on MRI images is the presence of non-brain tissues, eyeballs, and skin. Therefore, these parts should be stripped to obtain a clear image that can be used for further exploration. There are several skulls stripping methods, such as morphology-based, intensity-based, deformable intensity-based, hybrid-based, and atlas-based [22]. Further, there are software tools such as 3D slicer [23] for skull stripping manually other than the filter algorithms implemented in ITK [24]. Skull stripping follows the image registration, warping, and intensity normalization, as image preprocessing steps, in order to similarize the metadata of the sequences. Lao *et al.* [25] have used the open-source software ITK for skull stripping, rigid registration and intensity normalization via histogram matching. Nonetheless, Wijethilake *et al.* [26], [27] have used the open-source software 3D slicer for skull stripping and registration. Moreover, Tixier *et al.* [28] have implemented a C++ wrapper function with Insight ToolKit for the rigid registration of MRI for their work. In Table 1, we have summarized the commonly used tools available for structural MRI analysis of the brain.

3) SEGMENTATION

After image pre-processing, the aforementioned MRI sequences are used to delineate the regions of interest of the brain tumor into the pathological sub-regions known as necrosis, enhancement, and edema [34]. This can be done using both manual annotations and also automated techniques. ITK-SNAP [29] is the widely used software tool for manual segmentation. However, manual segmentation

TABLE 2. Segmentation methods and imaging features extracted in glioma related studies.

Related Work	Num. of patients	Segmentation	Features extracted
[44]	285	Ensemble model with CNN [50], DFKZ Net [38] & 3D U-Net [51]	First-order statistics, shape and texture features
[46]	285	An ensemble of 3D U-Nets [51]	6 volume features, age
[52]	285	VGG-16 [48] based FCN	Age, size and coordinates
[14]	163	ROI was manually annotated.	Texture, shape and volume features
[53]	163	ROI was manually annotated.	Histogram features
[54]	163	Deep U-Net [55]	Texture, volume, Euler, histogram
[47]	128	ROI was manually annotated.	Handcrafted texture features & deep features
[56]	69	ROI was manually annotated	clinical information & deep features
[39]	335	3D Attention U-Net	geometry, intensity & volume features
[41]	285	encoder-decoder based 3D architecture	-
[40]	369	3D U-Net architecture	-
[42]	335	Two-stage cascaded U-Net	-
[25]	112	ROI was manually annotated.	Geometry, intensity, texture features & deep features

is a tedious, laborious and time-consuming process. The annotation can also vary with the observer. Therefore, with the advancements in the field of artificial intelligence, DL based automated segmentation extensively utilizes well-known convolutional neural network (CNN) architectures such as U-Net [35], [36] and fully convolutional networks (FCN) [37]. The U-Net [35] architecture is capable of extracting both semantic and spatial information through the encoder-decoder sections while capturing fine details through the skip connections. Recently, 2D U-Net and as well as 3D U-Net, with minor modifications, have also shown promising improvements in glioma segmentation [38]. The attention block is a similar modifying state-of-the-art DL module integrated to U-Net and other encoder-decoder architectures for brain tumor segmentation [39]. In addition, skip connections are also integrated within the attention module, to avoid inconsistencies occur due to the fusion of spatial-wise and channel-wise features. Further, U-Net model, generative adversarial networks with different optimizers together with probabilistic programming languages have successfully used to segment medical images [36]. More recently, the DL researchers argue that the attention should be directed towards other factors that are important in optimizing DL networks, such as the per-sample and population loss functions and the optimizer, besides the DL architectures in order to obtain a finer brain tumor segmentation [40].

Myronenko [41] has proposed a model with high performances, which ranked 1st in the Brain Tumor Segmentation (BraTS) 2018 challenge, for brain tumor segmentation task. That study exploits a novel encoder-decoder based 3D architecture, with a large encoder and a small decoder, to extract image features and build the segmentation, respectively. They have added an extra decoder branch for the image reconstruction, where it is used to regularize the typical encoder, providing the ability to learn with limited datasets. Medical Image Computing and Computer Assisted Interventions (MICCAI) BraTS 2019 challenge 1st ranked winner, Jiang *et al.* [42] have proposed a two-stage cascaded U-Net, where the second stage consists of two decoders to boost the performance.

However, these automated segmentations do not accurately separate the brain tumor and the normal brain tissue due to the ambiguous boundaries and intensity variations. Hence, there exists room for enhancements in segmentation with DL.

4) FEATURE EXTRACTION

The MRI segmentation is followed by feature extraction. The texture features, intensity features, shape-based features are extracted from each subregion [43]. Mean, variance, skewness related to pixel intensities can also be extracted for a particular tumor sub-region. Further, shape features, such as area, centroid coordinates, length of axis, fractal dimensions, that depict the complexity of the region can be procured. Texture based features such as entropy, energy, correlation, dissimilarity are extracted from each subregion [15]. The features extracted from radiographic medical images are known as radiomics. Recently, deep features are also obtained from the MRI images; high-level features that depicts features unachievable with traditional feature extraction tools.

In Table 2, we have summarized related studies and the segmentation approaches with the extracted radiomic features for each work. Sun *et al.* [44] have extracted features based on grey level intensity, such as mean, median, standard deviation, variance, maximum and minimum intensity values, energy, and entropy values for edema, non-enhancing region, necrosis, and whole tumor regions. Moreover, for shape features, they have included the major and minor axis length, surface area to volume ratio, surface area, volume, and other features, that interprets the shape of each sub-region. The features that characterize the texture of each subregion, such as gray-level co-occurrence matrix features, gray-level run length matrix features, are extracted as well. Pyradiomics toolbox [45] has been used for feature extraction.

Sanghani *et al.* [14] have obtained 2D shape features, such as bounding ellipsoid volume ratio and orientation, spherical disproportion, and sphericity along with other 2D shape features and texture features. Alternatively, Feng *et al.* [46] have extracted volumetric features for each sub-region and integrated with resection status and age for survival prediction.

Han *et al.* [47] have extracted handcrafted radiomic features manually from the Region of Interest (RoI) and integrated with deep features obtained from Visual Geometry Group 19 weight layers(VGG-19) [48] model. Lao *et al.* [25]

have proposed a similar approach for survival prediction using CNN [49] to extract deep features. In that study, they also have extracted handcrafted features, that include geometry, texture, and intensity features obtained using Matlab in house feature extracting program.

C. GENOMICS

Gliomas involve genetic and epigenetic changes that can cause activation in the oncogenes and the inactivation in the tumor suppressor genes. The technological growth in the field of high throughput sequencing technologies has opened many opportunities in survival analysis. Thus, prognosis analysing studies related to glioma has manipulated different types of molecular profiles, including gene expression profiles, methylation profiles, and mutation profiles, that reflect those genetic and epigenetic alterations. In the meantime, the cost of acquiring those profiles has also decreased. As a result, a large number of glioma patients' genomic profiles are provided publicly as in The Cancer Genome Atlas (TCGA)[1] and Chinese Glioma Genome Atlas (CGGA).[2]

1) GENE EXPRESSION PROFILING

For genomic expression level analysis, microarray technology was used initially for obtaining the DNA sequences, with Agilent, Affymetrix, and Illumina platforms. The gene expression levels of the fragments of DNA are determined by comparing them with a standard target sequence. Yet, this technology is unable to identify pre-unknown genes. Solving this drawback, next-generation sequencing (NGS) came into the field of sequencing. Illumina HiSeq is a renowned modern platform in NGS, used for DNA and RNA sequencing. NGS can provide a more comprehensive view of the transcriptome by detecting pre-undetermined transcripts and noncoding regions, and thus, appropriate for gene expression profile extraction [57]. Further, these high throughput sequencing technologies such as NGS can accelerate modern genomic and proteomic research with massive volumes of data they generate [58].

2) METHYLATION PROFILING

Methylation is an epigenetic variation that adds a methyl group to the CpG sites. This can regulate the transcription of genes by acting on the particular genes promoter region, differentiating the expression of the corresponding gene. Illumina platforms, known as Infinium Human Methylation 450k (IHM-450k) and Infinium Human Methylation 27k, are frequently availed BeadChips kits for acquiring high throughput DNA methylation profiles [59].

3) MUTATION PROFILING

In addition, mutations in cells can cause tumorigenesis. Mutations can occur in tumor suppressor genes that repair DNA alteration and cell division, and hence, causing uncontrollable cell division. IDH1/IDH2 mutations are identified as a prominent mutation that occurs in glioma patients, which increase survival [60]. Mutation profiling identifies the nonsense, missense, and silent mutations from the obtained RNA sequences.

In survival prediction with genomics, gene expression profiles are frequently deployed for risk estimation of glioma patients [61], [62]. Nevertheless, associations between promoter methylation and gene expression are used for prognostic gene identification for risk estimations [63].

4) RADIOGENOMICS

Radiogenomics is an emerging field for survival analysis of many cancer related studies. In radiogenomics, the associations between radiomics and genomics are assessed.

Identifying imaging biomarkers that reflect the genomic behavior is very much convenient for the clinicians and other research personnel, as obtaining genomic profiles is an invasive and tedious task and also not been often used in clinical practice. Hence, GBM subtypes that have been divided based on molecular level alterations are recognized and predicted using the imaging biomarkers [27].

For survival prediction, radiogenomics are used in a couple of studies related to glioma. Wijethilake *et al.* [26] radiomics and gene expression features are fused together for survival prediction into short, medium, and long survival class prediction leading for higher accuracy. Tixier *et al.* [28] predict the survival by combining texture, shape features extracted from segmented tumor regions as radiomics, and the O-6-methylguanine-DNA methyltransferase (MGMT) methylation, IDH1 mutation status as genomics.

D. OTHER DATA TYPES USED IN SURVIVAL ANALYSIS

1) HISTOPATHOLOGY

The pathological images are microscopic images obtained from the tissue specimens of the glioma acquired with biopsy or surgery. Histopathologists engage in analysing microscopic images for diagnosis and prognosis purposes in clinical practice. As we mentioned in the introduction, the gliomas are divided into astrocytomas, ependymomas, and oligodendrogliomas based on this underlying histopathology. This histopathology behind gliomas is also used for survival estimation of gliomas in a couple of studies. Mobadersany *et al.* [64], Rathore *et al.* [65] have proposed a DL based approach for survival prediction of glioma patients with pathological images and integrates histology with gene biomarkers to predict the outcome.

2) CLINICAL INFORMATION

Most of the studies that use the above discussed data types have considered age as a clinical feature. The MICCAI BraTS challenge 2018 best performing work has also proved that age alone can predict survival with a high accuracy [52].

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

TABLE 3. Frequently used Glioma cohorts for survival analysis of Glioma patients.

Dataset	Data types					Studies that used the dataset
	Radiology	Pathology	Geno/Transcript-omics	Proteomics	Clinical	
TCGA	X	X	X		X	[26], [28], [61], [62], [66]–[68]
CPTAC	X	X	X	X	X	[69]
BraTS	X				X	[14], [15], [39], [44], [46], [52]–[54]
OBTS			X	X	X	[68]
CGGA			X		X	[61]
GEO			X		X	[62], [66], [67]

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

E. PUBLIC GLIOMA COHORTS

There are several publicly available datasets with imaging, genomic and histopathological records of glioma patients with their corresponding clinical information. The National Cancer Institute's (NCI's) Genomic Data Commons (GDC)[3] is a platform that provides access to Genetic, pathological and clinical data of The Cancer Genome Atlas (TCGA) datasets, including both GBM and LGG data collections. The Cancer Imaging Archive (TCIA) stores the radiological and pathological data of the correlative TCGA cases. Other than that, TCIA also provides access to new ongoing data collecting projects such as the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC), that also collect proteomics profiles of glioma patients. In addition, the Multimodal BraTS [4] provide pre-operative multimodal MRI scans of GBM and LGG patients along with clinical information. Unlike other glioma imaging cohorts, BraTS dataset also provides manual annotations of the glioma, thus, enabling it to be used as a benchmark dataset for glioma segmentation tasks.

The Chinese Glioma Genome Atlas (CGGA) is a database that consists of genomic data of Chinese Glioma patients. This includes whole-exome sequencing, mRNA sequencing, mRNA microarray, microRNA microarray, and DNA methylation profiles for GBM and LGG cases. Further, there are specific tools and applications that are developed to analyze the DNA mutation landscape, mRNA/microRNA expression profile, and DNA methylation profiles. Similar publicly available genomic data cohorts are Gene Expression Omnibus (GEO) [5] and the Ohio Brain Tumor Study (OBTS).

Summary of these frequently utilized datasets and the studies that have used them, are given in Table 3.

III. PREPROCESSING OF GLIOMA SURVIVAL RELATED DATA

1) DATA CLEANING

Data cleaning is a vital task in survival prediction to eliminate noise and improve the quality of data. Duplicate entries, outliers, and missing data are some of the facts that might reduce the quality of data. In medical information, missing data is a frequent problem. Generally, the records with missing data are removed or assigned values by calculating the mean or mode. The mean of the available continuous feature values is assigned to the missing data of the corresponding feature. If the feature values are categorical, the mode of the values is assigned to the missing data of the particular feature. However, according to the literature, removing the data will reduce the statistical power, and the imputations can affect the variance of the features [70].

2) NORMALIZATION

Normalization of features is mandatory to obtain better fitting data and model. Standardization, also known as Z score transformation, is a popular method 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 [18].

3) DIMENSIONALITY REDUCTION

Dimensionality reduction is essential in the medical field to optimize the performance of the learning algorithm by increasing the accuracy. According to Ladha and Deepa [71], 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 resources and 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 by 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

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

TABLE 4. Feature selection and analyzing techniques used in survival prediction of gliomas with radiomics.

Related Work	Feature selection				ML Technique						Accuracy
	IR	PCA	RFE	CC	RF	LR	SVM	XGB	ST	ANN	
[44]	x				x						61%
[46]						x					61%
[52]		x			x						61%
[14]			x				x				89%(CV)
[53]				x					x		67%(CV)
[54]			x					x			73%(CV)
[56]		x					x				89%(CV)
[15]										x	46.8%
[39]			x					x			48.3%

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; CV: Coefficient of Variation

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, the correlation has been utilized in feature selection with high dimensional genomic data [72]. 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.* [25] 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 [66].

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) [73] 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) [14], [54]. 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.

  Principal 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 [74].

IV. GLIOMA SURVIVAL ANALYSIS APPROACHES

Survival analysis of glioma patients is performed using various analysis methods, including ML, DL and statistical analysis based methods. Since early 2000, 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 features. Other than that, there are statistical approaches, prognostic risk score models, and prognostic nomograms proposed in glioma survival-related studies. In Table 4 and Table 5, we report feature extraction methods and the analysis methods followed by several studies that have used radiomics and radiogenomics, respectively as the input.

TABLE 5. Feature selection and analysing techniques used in survival prediction of gliomas with radiogenomics.

Related Work	Feature selection		Analysing Technique			Accuracy
	RFE	LASSO	KM	LR	SVM	
[28]		x	x			–
[26]	x			x	x	90%

RFE: Recursive Feature Elimination; LASSO: Least Absolute Shrinkage and Selection Operator; KM: Keplen Meier; LR: Linear Regression; SVM: Support Vector Machine

A. MACHINE LEARNING BASED SURVIVAL ANALYSIS

Different ML algorithms are utilized to predict the survival of glioma patients with a selected or filtered set of features. Since early 2000, many ML techniques have been used to predict the overall survival of glioma patient using radiomics, genomics and radiogenomics. 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.

1) SUPPORT VECTOR MACHINE

SVM is a frequently used supervised learning approach for classification and regression tasks in the domain of cancer research [75], [76]. 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. For instance, SVMs have successfully applied to analyse gene expressions [76]. 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.

$$\hat{\Phi}\left(\boldsymbol{x}'\right) = \text{sign}\left(\left\langle \mathbf{w}.\boldsymbol{x}'\right\rangle + b\right) \tag{1}$$

where, $\boldsymbol{x}'$, the test point, $\mathbf{w}$, the coordinates of the separating hyperplane and b, bias to the origin. SVM classifies 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 [14], [56]. Nie *et al.* [56] have obtained 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.* [14] have used a set of novel shape features that have a high contribution to the survival class prediction after selecting based on RFE. They have obtained a high accuracy proving that RFE, along with SVM, is an effective approach for survival prediction.

Emblem *et al.* [77] have presented 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 demonstrated that the proposed SVM based model can predict survival with higher accuracy compared to an expert in the clinical field. SVM has also been used for GBM subtype classification [27], [78].

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 and Zubek [79] has proposed a SVR algorithm for censored data by modifying the epsilon insensitive loss function asymmetrically.

2) 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 can 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, a 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.* [80] and Ciampi *et al.* [81]. 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 and Olshen [82] have initiated this with the Wasserstein metric between the survival functions, paving the ground for the criterion, based on the dissimilarity between the nodes. Ciampi *et al.* [83] have proposed log-rank statistics, likelihood ratio statistic, and Wilcoxon–Gehan statistic to measure the heterogeneity between the nodes. This method [84] has been utilized for establishing 6 prognostic groups of malignant glioma patients [85]. Moreover, this recursive partitioning analysis approach is being used for the overall survival analysis of lower-grade glioma patients based on pre-treatment factors [86]. With all these improvements, recently censored observations are restored with expected survival times, and the median survival tree is developed with L1 loss function [87].

The recent survival studies of glioma patients exploit survival trees with various types of data related to prognosis. Gandia *et al.* [88] have assessed 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 into 3 branches, where the first node criterion is the myo-inositol level and suggests that high myo-inositol levels can cause longer survival. The second branch cleaves the patients with a low myo-inositol 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 3 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.

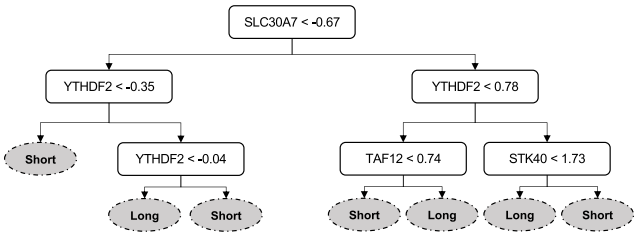

FIGURE 3. Sample survival tree for survival group classification into short, medium, and long survival.

3) 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 [89] and random forest regression [90] are proposed to overcome the instability by reducing the variance of simple trees.

In the Bagging algorithm, multiple versions of the bootstrap sample [91] are 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 [92]. This idea is extended by incorporating multiple cutting points for each selected feature and comparing them [93].

Recently, random forest regression has been widely utilized in radiomic based overall survival prediction [44], [52], as a regressor for overall survival prediction of glioma patients in days. Sun *et al.* [44] have achieved an accuracy of 61% with the random forest and have considered a limited number of samples might have caused the inadequate efficacy.

Choi *et al.* [94] 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.* [52] have derived 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.

4) 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 [95] have 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 [96]. Later, this gradient boosting algorithm is also modified for censored data in survival analysis [97].

5) LINEAR REGRESSION

Linear regression is a frequently used method for survival data analysis [26]. 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 [18].

6) 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 prior knowledge to make subtle inference and predictions. Bayesian approaches are also flexible by providing flexible algorithms, characterising the latent structure and uncertainty. 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 a particular event that occurred in between. The Bayesian posterior probability for a given feature set $\mathcal{F}$, and the model parameter $\mathbf{\Omega}$, is given by,

$$p(\mathbf{\Omega}|\mathcal{F}) = \frac{p_0(\mathbf{\Omega})p(\mathcal{F}|\mathbf{\Omega})}{p(\mathcal{F})} \tag{2}$$

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 $\mathbf{x}$ [98],

$$\begin{aligned} p(\mathbf{x}|\mathcal{F}) &= \int p(\mathbf{x}, \mathbf{\Omega}|\mathcal{F})\mathrm{d}\mathbf{\Omega} \\ &= \int p(\mathbf{x}|\mathbf{\Omega}, \mathcal{F})p(\mathbf{\Omega}|\mathcal{F})\mathrm{d}\mathbf{\Omega} \end{aligned} \tag{3}$$

There are two Bayesian methods, that are being widely used in the glioma survival prediction [99]–[101].

- Naïve Bayes (NB) [102] Based on the Bayesian rule (equation 2), for a given feature set ($\mathbf{x}$ the probability of belonging into class ϕ, 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 probabilities ($P(\phi_1|\mathbf{x})$ 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 ϕ_1 and if $P(\phi_1|\mathbf{x}) < P(\phi_2|\mathbf{x})$ then $\mathbf{x}$ belongs to class ϕ_2.

The class is chosen arbitrarily if the posterior probabilities are equal. However, this Naive Bayes method assume a mutual 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 [103], [104]. Nonetheless, as a white-box this is easily understandable for clinicians and other personnel.

- Bayesian networks (BN) [105] 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.* [17] for overall 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.* [100], [106] have leveraged 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 and Frey [107] for predicting survival status after 2 years from the diagnosis with molecular-level data.

7) ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANN) are ML algorithms that resemble the biological neural systems, which were established by Rosenblatt in 1958 [108]. Thus, based on the neuronal activities, the concept of ANN was initially established by McCulloch and Pitts [109]. A basic ANN model, as shown in Fig 4, 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 delivers 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.

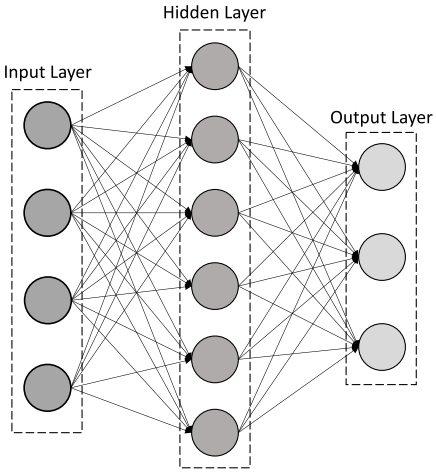

FIGURE 4. Artificial Neural Network consists of a single hidden layer.

In the review of the literature, Faraggi and Simon [110] developed an ANN for survival analysis of cancer for the first time in 1995. Recently, neural network is incorporated with the Cox proportional-hazards 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 [111]. Neural Network is utilized in glioma survival prediction to predict the survival time in days directly. Islam *et al.* [15] have proposed 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.

B. DEEP LEARNING BASED SURVIVAL ANALYSIS

Recently, DL has drawn the attention of 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 IV-A. 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.* [112] have proposed a deep neural network, followed by a cox survival model for outcome prediction with high dimensional genomic data. They have used deep survival neural network as the DL algorithm. As an extension of this study, Mobadersany *et al.* [64] have integrated 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 is 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.

Further, Rathore *et al.* [65] have proposed an overall survival classification approach with the Residual Network (ResNet) architecture for pathology images. They have shown an accuracy of 84.32% for survival class prediction, long and short survival prediction in glioma patients. Both of these studies have claimed the need for automated patch extraction methods, to overcome the burdens, such as the time that occur with manual patch extraction.

Several related studies have used the TCGA dataset. However, a major concern associated with training a DL network is the requirement of a large amount of labeled data. This occurs as a result of a high number of (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 a specific set of data augmentation techniques followed by DL researchers in the 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, the sampling regions of interests (patches) from the pathological images allow the DL models to discern the variations effectively. This heterogeneity is visible in pathological images, and thus, Mobadersany *et al.* [64] have selected the regions of interest from each slide with the expert knowledge.

C. STATISTICAL ANALYSIS BASED SURVIVAL ANALYSIS

There are three types of statistical analysis methods for survival analysis. *i*) Non-parametric *ii*) Semi-parametric and *iii*) Parametric [18]. 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.

1) 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. Accordingly, if an event, such as death, occurs during the monitoring, 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 [18].

2) KAPLAN MEIER ANALYSIS

In survival analysis, there are two functions depend on the time. *i*) Survival function: Probability of surviving at least up to time t; $(\text{pr}(T \geqslant t))$. *ii*) Hazard function: Conditional probability of dying at time t, if the patient survived until time t. Keplan-Meier Curve [113] 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.

$$\text{S}(\text{t}') = \text{p}_1 \times \text{p}_2 \times \text{p}_3 \times \ldots \times \text{p}_{\text{t}'} \tag{4}$$

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

$$p_i = \frac{k_i - d_i}{k_i} \tag{5}$$

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.

3) 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 the two survival curves (two population groups).

$$\chi^2(\log \text{rank}) = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \tag{6}$$

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. Although, 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 [18].

4) COX PROPORTIONAL HAZARD ANALYSIS

Cox proportional hazard model [114] is semi-parametric survival analysis approach and a multiple regression model, that can account for many factors at once for the analysis. Hazard function $\lambda(t)$, the probability of dying at a particular time if survived up to that time, is the dependent variable of this model.

$$\lambda(t) = \lim_{\Delta t \to 0+} \frac{\text{pr}(t \leqslant T < t + \Delta t | t \leqslant T)}{\Delta t} \tag{7}$$

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

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\mathbf{z}\beta) \tag{8}$$

TABLE 6. Related work on prognostic risk score calculation.

Related Work	Glioma type	Datasets	Feature selection method	Features used
[61]	GBM	TCGA & CGGA	Cox PH	Gene Expression data
[62]	Glioma	GEO & TCGA	Cox PH	Gene Expression data
[63]	LGG	-	Cox PH	Gene Expression data & Methylation data
[66]	GBM	TCGA & GEO	log-rank test	Methylation data
[67]	GBM	TCGA & GEO	Cox PH	Gene Expression data
[69]	GBM	CPTAC & TCIA	Cox PH	Radiomic features

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

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

$$\frac{\lambda(t, X_1)}{\lambda(t, X_2)} = \frac{\lambda_0(t)\exp(X_1\beta)}{\lambda_0(t)\exp(X_2\beta)} = \exp[(X_1 - X_2)\beta] \quad (9)$$

This demonstrates that the 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 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.* [61] have selected 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.* [115] have also utilized a univariate cox model for identifying survival-related genes.

D. OTHER METHODS USED IN SURVIVAL PREDICTION

1) RISK SCORE MODEL

Risk score modelling is another common approach for risk estimation of glioma patients. Thus, the patients are separated into the high risk and low-risk groups based on the expression or methylation of genes or other features associated with the survival. In order to obtain the most prominent genes associated with survival, univariate and multivariate cox proportional hazard analysis are frequently utilized [61]–[63]. 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, using the following formula.

$$\text{Risk score} = \sum \text{status}_{gene} * \text{coef}_{gene} \quad (10)$$

where coeff_{gene} is the coefficient obtained from the Cox PH hazard analysis for the corresponding gene and 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.* [61] 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 these 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.* [62] 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.* [69] extract radiomic features from T1 weighted MRI to develop risk score models, separately based on the gender for accurate risk estimation. In Table 6, we present the summarize of several related works that have developed prognostic risk score models.

2) 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.* [116] have developed a 5 gene prognostic nomogram for GBM patients, considering the gene expression levels of OSMR, BICDL1, SH3BP2, MSTN, and RGS14 genes. Similarly, Gittleman *et al.* [68] have proposed 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.* [117] to develop a comprehensive nomogram for glioma patients.

An example of a prognostic nomogram is shown in Figure 5. 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.

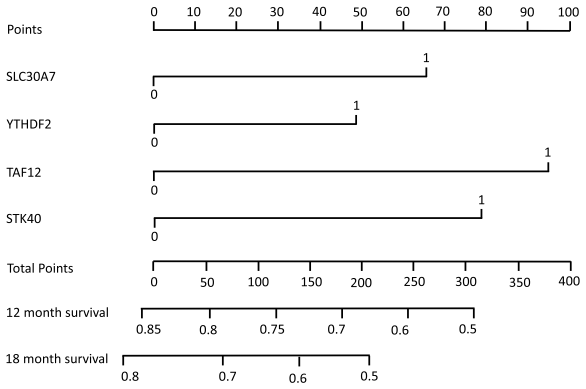

FIGURE 5. A sample nomogram was developed to estimate the probability of survival after 12 months and 18 months from the diagnosis. The expression levels of SLC30A7, UTHDF2, TAF12, and STK40 genes are considered in this.

A summary of related works that have proposed prognostic nomograms for survival probability estimation is given in Table 7.

TABLE 7. Related work on prognostic nomograms development.

Related Work	Glioma type	Datasets	Features used
[116]	GBM	TCGA & GEO	Gene Expression data
[68]	LGG	OBTS & TCGA	Clinical & molecular factors
[117]	Glioma	private dataset	Hematological & clinicopathological factors

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

V. DISCUSSION

Survival Prediction of glioma patients is a critical challenge for oncologists and other clinicians as it directly influences the patient in many aspects. Over the past decades, many research personnel has focused on developing survival prediction approaches by utilizing various types of data and techniques.

A. IMPORTANCE AND IMPACT OF GENOMICS IN PROGNOSIS

Many studies have been conducted to identify the importance of genomics in gliomas, and also, some genes have shown associations with survival. The following molecular alterations are the most prominent alterations that are found to have prognostic significance in glioma patients.

1) IDH1 MUTATION AND ITS IMPACT ON PROGNOSIS

IDH1 catalyze the oxidative decarboxylation of isocitrate to α-ketoglutarate as [118]. This plays a vital role in cellular protection from oxidative stress through the production of NADPH [119]. WHO grade I astrocytomas were not identified with mutations in IDH1 or IDH2. In addition, WHO grade II & III gliomas, i.e. LGG, mostly carry mutations in IDH1, including the secondary WHO grade IV - GBM, which arise from lower-grade gliomas [120].

In spite of all the studies, Yan *et al.* [120] have further determined that the GBM patients with IDH1 mutations have a median overall survival of 31 months, where IDH1 non-mutant patients have a median survival of 15 months, significantly lesser than the IDH1 mutant patients. Moreover, patients with anaplastic astrocytomas have a median survival of 65 months and 20 months for IDH1 mutations and non-mutations, respectively [120]. In fact, IDH1 mutation is considered as a positive prognosis marker in lower-grade gliomas [121]. Thus, researches verify that IDH1 mutation as the most important prognostic marker despite the underlying histology [60].

2) TP53 MUTATION

TP53 functions as a tumor suppressor protein by regulating cell division, which prevents cell division and growing too fast or in an uncontrolled way. If a damaged DNA cannot be repaired, the resultant protein of TP53 inhibits DNA replication. Thus, mutated TP53 is unable to inhibit the replication of genetically unstable cancerous cells [122]. Thus, mutant TP53 provides an independent pathway leading to gliomagenesis [123].

Many studies have revealed the association between the poor prognosis of lower-grade gliomas and TP53 mutation, and further analysis demonstrates that TP53 mutation is an adverse prognostic marker for patients with astrocytic and oligoastrocytoma histology [124]. Nevertheless, TP53 mutation and expression of mutant TP53 in GBM patients show an inverse correlation with the prognosis, and the chemosensitivity for temozolomide also reduce with the mutant TP53 [125]. However, researches have further demonstrated that the prognostic effects of TP53 in GBM patients depend on the clinical factor, age [126].

3) TERT PROMOTER MUTATION

TERT encodes telomerase, which maintains the length of the telomere region in the eukaryotic chromosome. Telomere and telomerase together play a crucial role in both tumors suppressing and tumor promotion [127]. Consequently, TERT promoter mutation is associated with increased expression of the TERT gene [128]. Therefore, tumor promotions occur by maintaining the telomere length through the overexpression of the TERT gene [129]. Thus, TERT promoter mutation occurs frequently in gliomas [130], [131]. Mutations in TERT

promoters are found in 58% of primary GBM and less frequently (28%) in secondary GBM [132]. TERT mutation is also common in patients with Oligodendrogliomas [133].

Nevertheless, TERT promoter mutation shows a low survival rate in GBM patients [128], [132]. Further, recent studies have shown that TERT promoter mutation has a positive correlation with age and the WHO grade [134]. Studies have revealed that TERT promoter mutation, along with a high relative telomere length, not only affects the survival of glioma patients but also determine the resistance to radiotherapy [135].

4) MGMT METHYLATION

MGMT encodes a protein that repairs DNA, damaged by alkylating agents [136]. Thus, MGMT unmethylated gliomas are resistant to alkylating chemotherapeutic agents, causing shorter survival of patients [137]. Moreover, methylation of MGMT silence the gene causing high sensitivity for alkylating therapeutic agents and, thus, for a better survival [138]. MGMT promoter methylated GBM patients treated with temozolomide have a better prognosis (median survival 36 months) than the patients without methylated MGMT promoter (median survival 16.8 months) [28].

5) GENE EXPRESSION PROFILE ANALYSIS FOR GLIOMA PROGNOSIS

The aforementioned different genetic aberrations in gliomas and other environmental effects, cell division, genetic inherent disrupt the cell behavior in cells. Thus, the regular cell behavior changes and its protein production, i.e., translations of genes get affected. These damaged cells reproduce rapidly and turn into cancers. Therefore, analyzing gene expression profiles exploits associations with molecular level genetic alterations [125], [128]. In fact, genetic profiling reveals risk factors, including protective traits such as immune responses, allergies.

Typically, GBM is the most heterogeneous glioma, which makes it challenging to treat patients. Overcoming this heterogeneity, gene expression profiles classify GBM patients into 4 subtypes, based on the molecular pathogenesis, illustrating the prognosis of each subtype [139]. Gene expression profile analysis, along with mutation profiles, are also used to develop targeted therapies and other personalized drugs for glioma patients, which also increase the overall survival [7].

In overall, genomics have a clear association with the occurrence, progression and prognosis of gliomas. Thereby, we speculate the importance of exploring genomics for the glioma survival prediction, to develop more stable and accurate platforms.

B. IMPORTANCE AND IMPACT OF RADIOMICS IN PROGNOSIS

Radiomics have indicated intuitions to genomic and clinical features of Glioma patients through radiomic analysis. Thus, radiomics are being used to predict genetic alterations such as IDH1 status, gene expression levels, and most frequently the survival. Unlike genomics, radiomics provide a non-invasive and low-cost approach for feature extraction and analysis. Therefore, radiomics are used to predict the molecular subtypes associated with the genomics without performing any surgical intervention. Although genomics has a clear insight on survival or the categorization of gliomas, the ability in radiomics to predict the clinical parameters makes it easy for the clinical practitioners to decide treatments due to the non-invasive behavior.

Nevertheless, the genetic information obtained from the tumor region might vary within the tumor. This is commonly known as intra-tumor heterogeneity, which occurs at molecular and histopathological levels. This limitation in related genomic analysis can be avoided when using radiomics for the analysis by considering tumor regions comprehensively, allowing spatial mapping of distinct molecular level changes. Moreover, based on these radiomics, 3 subtypes, known as enhancing, irregular and solid, are identified, with each subtype comprising similar molecular alterations.

Many studies have found associations between imaging features and the prognosis of glioma patients. Hammoud *et al.* [34] had identified that glioma patients with less or no necrosis and enhancement regions tend to survive longer compared to the patients with greater volume of those regions.

C. ADVANTAGES OF LEARNING APPROACHES TOWARDS SURVIVAL ANALYSIS

Learning approaches can learn from the large amounts of data, without explicitly specifying the prior assumptions, rules and limits by humans. However, traditional statistical methods are often based on assumptions such as the additivity of the parameters in proportional hazard models and linear functions [140]. Rajula *et al.* [140] have further showed that there are conflicts in the assumptions of proportional hazard method used in survival analysis studies in gastric cancer.

Learning approaches can handle a large number of predictors where a small number of observations are available. For example, using genomics, where thousands of gene profiles are available, and only some observations, such as long vs short survival, can be modelled using learning approaches straightforwardly. Traditional statistical methods have to deal with the limitation of handling only a limited number of choices of factors or predictors.

Nonetheless, learning algorithms can analyse various distinct data types, such as genomics, images at once, for predicting outcomes [26]. Also, from a clinicians perspective, statistical approaches are not straightforward and implemented in clinical practice is questionable and non-viable. Also, the interpretability of the learning methods is yet ambiguous. However, recently, the Explainable Artificial Intelligence has drawn the attention of the medical data science researchers, resolving the interpretability issue in the medical community as we discuss in Section V-E1. These constraints associated with explainability and interpretability of the learning approaches will be addressed in the future.

Among the learning approaches, DL methods have shown more flexibility towards survival prediction. However, the data hunger in DL approaches makes most of the survival related studies to use only ML methods, as they do not have thousands of parameters to be tuned as in DL paradigms. However, this will also be solved in the coming decade with the availability of more patient profiles, since most of the institutes have now focused on collecting data. Nevertheless, there are emerging deep probabilistic learning approaches, that can overcome limitations such as uncertainty, extensibility in traditional DL methods. Integration of these novel techniques will lead the current survival analysis to a new era.

D. LIMITATIONS OF EXISTING STUDIES

1) LIMITATIONS OF SURVIVAL ANALYSIS WITH GENOMICS

- Heterogeneity
  Specific improvements in the medical field are required to integrate survival analysis, peculiarly with genomics, into clinical management. For instance, molecular testing and sequencing procedures should be standardized and cost-effective to be implemented in practice. In addition, it is a challenging task to develop common data standards due to the complexity of molecular-level data and the various high throughput platforms such as next-generation sequencing and microarray used to acquire different datasets. The technical limitations can cause the same gene expression profiles obtained from different platforms to be inconsistent. The microarray gene expression profiling can get affected by the cross-hybridization of the probes and the limited detection range of individual probes [141]. Hsu *et al.* [115] has identified this as a drawback for the identification of survival related genes, that makes their observed survival related genes differ from the other related studies. Wijethilake *et al.* [17] have also reported this as a limitation for testing the proposed model on a separate dataset and thus, use cross-validation to verify the performance of the proposed method for survival prediction.
  Nonetheless, in clinical practice obtaining genomic profiles requires the tissue samples obtained through invasive surgical resection or biopsy. However, the procured tissue sample, small biopsy specimens may not accurately represent the molecular pathogenesis of the entire tumor, which is mostly heterogeneous [4]. Due to this heterogeneous nature in gliomas Xian *et al.* [62] have claimed that a single biomarker is not sufficient for survival prediction and further explorations are required in identifying prognosis related biomarkers for clinical implications.
- Real-Time Acquisition
  Further, real-time sequencing of whole-genome exons, along with gene expression profiling, is a challenging task with the current technology. In clinical practice, genomics is not routinely used due to the high cost and the time.

2) LIMITATIONS OF SURVIVAL ANALYSIS WITH RADIOMICS

- Issues related to Image acquisition and feature extraction pipeline
  The datasets obtained from different institutions might have followed different hardware and acquisition protocols. This can also vary within the institutes. Therefore the images acquired are quite challenging to generalize. The radiomic feature extraction pipeline consists of several steps, discussed in Section II-B that are prone to errors. For instance, the manual annotations of the tumor regions can also vary with the annotator and thus, impact the features derived from the tumor regions. The handcrafted feature extraction formulas capture the textural and morphological attributes of a given region based on predefined mathematical formulas. However, these implementations also change with the software used and the techniques used. Most of these features are general features used in any kind of images, without specifying on medical images. Therefore, the implementation might require slight changes to deal with heterogeneous medical images obtained from different patient cases. Feng *et al.* [46] have mentioned a wide range of MRI protocols as a possible reason for the significant gap in segmentation performance between training and the testing phases.

3) COMMON LIMITATIONS IN SURVIVAL ANALYSIS

- Lack of Data
  Most of the studies have used publicly available, open-access datasets such as The Cancer Genome Atlas (TCGA), Chinese Glioma Genome Atlas (CGGA). Nevertheless, these cohorts also consist of less than 500 glioma patient cases, due to the limitations in time and cost in data acquisition. Therefore in survival prediction tasks are more likely to overfit and as a solution, some studies only include most survival related features to avoid overfitting [46]. BraTS challenge confirms the ML algorithms ability to handle smaller datasets over DL approaches [16]. Besides, Mobadersany *et al.* [64] have also disclosed the need for large datasets in DL paradigms for future research. In addition, some of the studies also consider their own local datasets for the research work. This causes issues in replicating the same experiments in external cohorts, and also comparisons with other related studies become critical. The missing data in datasets also makes this limitation more critical. Some work they have neglected the patients' records with missing data or have dealt with this issue, as we mentioned in Section III-1.
- Data Scale and Computation requirements
  The data, genomics, and also images require large volumes and high-performance computing resources to store and process. However, despite the rapid growth of data volumes, the computing resources have a slow development, posing a significant challenge in handling extensive data. Nevertheless, the DL tasks in survival

analysis, such as automated segmentation, are computationally expensive to train despite the requirement of a massive amount of data for an accurate prediction. A major challenge that occurs when handling genomics is the ability to make plausible sense with large volumes of data. Integrating these large volumes, with different molecular interpretations, along with clinical data, is also challenging.

- Reproducibility for clinical application
  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 [142]. 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 [143]. Moreover, the reproducibility of deep features, extracted from imaging data, is also a critical task and, thus Han *et al.* [47] have incorporated 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, providing improvements for the field of precision medicine in cancer and genomic testing in practice [144].
  In contrast, clinical parameters have shown more promising improvements in survival prediction as they are generalizable between different institutions unlike other data, that are subjected to variations in acquisition. Thus, In BraTS 2018 challenge the best performing model has observed the strongest correlation between age and the survival, and further the third best performing survival prediction model has only utilized age for the survival prediction as reported by Bakas *et al.* [16], Feng *et al.* [46].
- Imbalanced Datasets
  Some of the survival-related analyses are performed as a classification task, as in short, medium, and long survival group-wise classification, or as an analysis study focusing on the censoring status. For the learning models to effectively learn all the classes, the number of patient cases in each class should be equal. Alternatively, the learning models are likely to give priority to the classes with more learning instances in survival prediction. There are commonly used sampling techniques to avoid the class imbalance in datasets by oversampling the minority class and under-sampling the majority class. Wijethilake *et al.* [17] have used minority class oversampling for overall survival prediction with genomics to avoid the class imbalance.
- Multidisciplinary requirements
  Research Personnel with technical, clinical knowledge and expertise are necessary to develop a survival analysis tools, efficiently utilizing the available resources. High technical people tend to incorporate novel technologies and facilities for these analyses, and the final outcome can be less relevant in clinical practice. On the contrary, clinical researchers are likely to consider more on clinical perspectives, in spite of the technical errors such as robustness, multiple comparison errors. Multidisciplinary research groups will avoid these obstacles and bring up technically efficient and accurate tools or models with considerable clinical importance.
  In addition, some of the discussed studies have got the experts supervision from a distinct disciplinary, for different tasks such as for manual segmentation of tumor regions [16] and for the region selection of pathology images [64]. This is also important for the development of useful survival analysis, with a coherent underlying concepts.

E. FUTURE RESEARCH DIRECTIONS

In order to compensate for the aforementioned limitations, the technological advancements, along with clinical applications, are progressed into novel advanced techniques. These directions can be identified as follows.

1) EXPLAINABLE ARTIFICIAL INTELLIGENCE

The black box behavior of Artificial Intelligence algorithms was questionable recently, with the strive to recognize the importance of most relevant and how the predictions are made. Moreover, in spite of the high accuracy, the clinicians are doubtful to accept a survival prediction tool blindly without a proper understanding. Several approaches are being proposed to overcome this black box behavior, explaining the underlying algorithm and the contribution of the features. SHapley Additive exPlainations (SHAP) [145] is an explainable technique used to delineate the ML models following a game-theoretic approach. Another renowned explainable method is Saliency mapping [146], where a heat map is visualized expressing the impact of each individual pixel of the prediction and thus, explaining the classification or segmentation neural network models. With this, the clinicians can make sure the segmentation is error-free prior to feature extraction and prediction.

2) AVAILABILITY OF BIOINFORMATICS TOOLS AND COMPUTATION ENHANCEMENTS

The growth in the available computational frameworks and libraries implemented in different programming languages such as R and python has made the prediction task more implementable at ease. Since most of these tools are also

open access, the researchers tend to apply various techniques and compare the performance, thus, analytically identifying the most suitable survival prediction technique. Moreover, the evolution in Graphical Processing Units (GPUs) has made the DL a possible task with parallel processing [58], [147]. In addition, there is a possible research direction to implement the classification algorithms with an in-build parallelism [148]. Further, DL techniques can combine with probabilistic models, Deep probabilistic programming (DPP), to increase the efficiency and flexibility of the computations [36].

3) PRECISION MEDICINE

Precision medicine analyzes the characteristics of the glioma considering the genetic alterations, imaging markers, prognosis of the patient to personalize the required treatments of each person. In this scenario, the survival estimation of glioma plays a crucial role in deciding the ideal treatment for each individual that will consequently help to improve the patient outcome. However, relying on genomics is genomics for personalized medicine is challenging due to tumor heterogeneity. Therefore biopsies are repeatedly done as the tumor progresses, and the personalized therapies are determined. For this, the initial overall survival or the risk estimation is necessary to manage the treatments efficiently and effectively.

4) ADVANCES IN GENOMIC TECHNOLOGIES

Next-generation sequencing has formed a conspicuous revolution in the high throughput technologies, revealing 'driver' gene alterations that occur in gliomas. The advancements in the high throughput technologies have reduced the cost of analyzing omics data, including genome, transcriptome, proteome, and metabolome. Further, these technologies have enabled gene expression, methylation, and mutation profiling of a large number of genes at once. This transcriptome analysis has led new directions in glioma prognosis analysis, despite clinical and other non-invasive imaging factors. However, as we discussed earlier, the intra-tumor heterogeneity is a critical constraint in survival analysis with genomics. Ultra-deep sequencing sheds light on this by sequencing the different regions of the same tumor [149]. Thus, in the future, more comprehensive studies can be performed integrating genomics, transcriptomics, and epigenomics, the omics beyond the exome.

5) DATA-PRIVATE COLLABORATIVE LEARNING METHODS

Most of the recent studies related to glioma prognosis have shown promising improvements in research with the availability of data collected by various institutions. Thus, each institutions tend to collect and analyse data by themselves. However, these DL based models have shown biases and inability to perform well with other unseen data acquired by other institutions. In the current exploration, some works and publicly available cohorts merge data from multiple institutions, despite the concerns associated with the data ownership, privacy and technical configurations. Data-private collaborative learning methods are used in learning paradigms recently, facilitating many institutions to collaborate their data without sharing them. This has several implementations such as federated learning, where the multiple institutions train the learning model parallelly with their data and aggregate learned models at a central server. Federated learning has shown promising improvements for Glioma segmentation [150]. This collaborative learning can also be serial, by updating the ML model and passing it for the next. Data-private collaborative learning can be used in future work related to survival driving the precision of prediction algorithms to a higher level, while maintaining common acquisition protocols, common preprocessing routines and similar data harmonization between institutions [151].

The genomic alterations associated with survival were investigated since the 1990s, and the utilization of genomics for survival or risk prediction is frequently seen after 2018. Admittedly, imaging data were widely used for survival prediction since early 2000. Yet, the advancements in segmentation and feature extraction algorithms are mostly witnessed after 2010. In this exploration, we have considered the most recent survival prediction approaches followed by various multidisciplinary research groups, in 2016-2020 period.

Furthermore, the future of the survival prediction for glioma patients is drifting towards the utilization of high dimensional inputs and models, whereas standardization and generalization is a pivotal point. Multi-institutional collaborations with data-private collaborative learning can drive this field more forward. The main obstacles that arise can be stated as the various acquisition, preprocessing protocols followed by different institutions. Even after these obstacles are solved, the clinical deployment of survival prediction should happen after close monitoring, testing and regularizing of the standards in an application.

VI. CONCLUSION

In study has focused on survival analysis approaches followed by existing studies related to gliomas, that has a low survival compared to other cancer types. In the modern era of personalized medicine, the survival prediction of glioma patients are performed utilizing diverse imaging and gene biomarkers. Although this area is being improved and drawn the attention of the clinicians in the past couple of decades, still, there are requirements for significant improvements to provide survival prediction routinely in clinical practice. Through this survey, we have explored the potential of genomics in glioma survival analysis, although it has an invasive approach. We further reported the importance of using imaging for survival prediction, due to its non-invasive behavior and as imaging is frequently used as an initial diagnostic tool. In this survey, we disclosed the limitations in both approaches and enlighten the researchers who are interested in survival prediction of glioma patients to follow novel strategies to create maximum insight out of miscellaneous data. We further emphasized the importance of clinicians'

feedback to develop clinically feasible survival prediction software application. Thereby the future glioma management will thrive with accurate survival prediction approaches, ultimately provoking a better prognosis in glioma patients.

ACKNOWLEDGMENT

The authors acknowledge the support received from the Conference & Publishing grant, University of Moratuwa, Sri Lanka for publishing this paper.

REFERENCES

[1] Q. T. Ostrom, H. Gittleman, G. Truitt, A. Boscia, C. Kruchko, and J. S. Barnholtz-Sloan, "CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2011–2015," *Neuro-Oncology*, vol. 20, no. 4, pp. iv1–iv86, 2018.

[2] Q. T. Ostrom, H. Gittleman, J. Fulop, M. Liu, R. Blanda, C. Kromer, Y. Wolinsky, C. Kruchko, and J. S. Barnholtz-Sloan, "CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the united states in 2008-2012," *Neuro-Oncology*, vol. 17, no. 4, pp. iv1–iv62, Oct. 2015.

[3] Q. T. Ostrom, H. Gittleman, P. Liao, C. Rouse, Y. Chen, J. Dowling, Y. Wolinsky, C. Kruchko, and J. Barnholtz-Sloan, "CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2007–2011," *Neuro-Oncology*, vol. 16, no. 4, pp. iv1–iv63, Oct. 2014.

[4] M. Weller, W. Wick, K. Aldape, M. Brada, M. Berger, S. M. Pfister, R. Nishikawa, M. Rosenthal, P. Y. Wen, R. Stupp, and G. Reifenberger, "Glioma," *Nature Rev. Disease Primers*, vol. 1, no. 1, p. 15017, 2015.

[5] G. N. Fuller and B. W. Scheithauer, "The 2007 revised world health organization (WHO) classification of tumours of the central nervous system: Newly codified entities," *Brain Pathol.*, vol. 17, no. 3, pp. 304–307, Jul. 2007.

[6] F. Sahm, D. Reuss, C. Koelsche, D. Capper, J. Schittenhelm, S. Heim, D. T. W. Jones, S. M. Pfister, C. Herold-Mende, W. Wick, W. Mueller, C. Hartmann, W. Paulus, and A. von Deimling, "Farewell to oligoastrocytoma: *In situ* molecular genetics favor classification as either oligodendroglioma or astrocytoma," *Acta Neuropathol.*, vol. 128, no. 4, pp. 551–559, 2014.

[7] K. K. Jain, "A critical overview of targeted therapies for glioblastoma," *Frontiers Oncol.*, vol. 8, p. 419, Oct. 2018.

[8] H. Ohgaki and P. Kleihues, "Epidemiology and etiology of gliomas," *Acta Neuropathol.*, vol. 109, no. 1, pp. 93–108, Jan. 2005.

[9] P. Yang, Y. Wang, X. Peng, G. You, W. Zhang, W. Yan, Z. Bao, Y. Wang, X. Qiu, and T. and Jiang, "Management and survival rates in patients with glioma in China (2004–2010): A retrospective study from a single-institution," *J. Neuro-Oncol.*, vol. 113, no. 2, pp. 259–266, 2013.

[10] E. B. Claus, K. M. Walsh, J. K. Wiencke, A. M. Molinaro, J. L. Wiemels, J. M. Schildkraut, M. L. Bondy, M. Berger, R. Jenkins, and M. Wrensch, "Survival and low-grade glioma: The emergence of genetic information," *Neurosurgical Focus*, vol. 38, no. 1, p. E6, Jan. 2015.

[11] M. Klein, "Neurocognitive functioning in adult WHO grade II gliomas: Impact of old and new treatment modalities," *Neuro-Oncology*, vol. 14, no. 4, pp. iv17–iv24, Sep. 2012.

[12] M. Moghtadaei, M. R. Hashemi Golpayegani, F. Almasganj, A. Etemadi, M. R. Akbari, and R. Malekzadeh, "Predicting the risk of squamous dysplasia and esophageal squamous cell carcinoma using minimum classification error method," *Comput. Biol. Med.*, vol. 45, pp. 51–57, Feb. 2014.

[13] M. S. Bal, V. K. Bodal, J. Kaur, M. Kaur, and S. Sharma, "Patterns of cancer: A study of 500 Punjabi patients," *Asian Pacific J. Cancer Prevention*, vol. 16, no. 12, pp. 5107–5110, 2015.

[14] P. Sanghani, B. T. Ang, N. K. K. King, and H. Ren, "Overall survival prediction in glioblastoma multiforme patients from volumetric, shape and texture features using machine learning," *Surgical Oncol.*, vol. 27, no. 4, pp. 709–714, Dec. 2018.

[15] M. Islam, V. J. M. Jose, and H. Ren, "Glioma prognosis: Segmentation of the tumor and survival prediction using shape, geometric and clinical information," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke Traumatic Brain Injuries*, A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, and T. van Walsum, Eds. Cham, Switzerland: Springer, 2019, pp. 142–153.

[16] U. Baid, S. U. Rane, S. Talbar, S. Gupta, M. H. Thakur, A. Moiyadi, and A. Mahajan, "Overall survival prediction in glioblastoma with radiomic features using machine learning," *Frontiers Comput. Neurosci.*, vol. 14, no. 61, pp. 1–9, Aug. 2020, doi: 10.3389/fncom.2020.00061.

[17] N. Wijethilake, D. Meedeniya, C. Chitraranjan, and I. Perera, "Survival prediction and risk estimation of glioma patients using mRNA expressions," in *Proc. IEEE 20th Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2020, pp. 35–42.

[18] P. Wang, Y. Li, and C. K. Reddy, "Machine learning for survival analysis: A survey," *ACM Comput. Surv.*, vol. 51, no. 6, pp. 1–36, 2019.

[19] L. Ganau, M. Paris, G. K. Ligarotti, and M. Ganau, "Management of gliomas: Overview of the latest technological advancements and related behavioral drawbacks," *Behavioural Neurol.*, vol. 2015, pp. 1–7, 2015.

[20] S. Metcalfe and R. Grant, "Biopsy versus resection for malignant glioma," *Cochrane Database Systematic Rev.*, vol. 1, no. 3, 2001, Art. no. CD002034.

[21] N. Upadhyay and A. D. Waldman, "Conventional MRI evaluation of gliomas," *Brit. J. Radiol.*, vol. 84, no. 2, pp. S107–S111, Dec. 2011.

[22] P. Kalavathi and V. S. Prasath, "Methods on skull stripping of MRI head scan images—A review," *J. Digit. Imag.*, vol. 29, no. 3, pp. 365–379, 2016.

[23] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper, and R. Kikinis, "3D slicer as an image computing platform for the quantitative imaging network," *Magn. Reson. Imag.*, vol. 30, no. 9, pp. 1323–1341, Nov. 2012.

[24] S. Bauer, T. Fejes, and M. Reyes, "A skull-stripping filter for ITK," *Insight J.*, vol. 2012, pp. 1–7, Jan. 2013.

[25] J. Lao, Y. Chen, Z.-C. Li, Q. Li, J. Zhang, J. Liu, and G. Zhai, "A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme," *Sci. Rep.*, vol. 7, no. 1, pp. 1–8, Dec. 2017.

[26] N. Wijethilake, M. Islam, and H. Ren, "Radiogenomics model for overall survival prediction of glioblastoma," *Med. Biol. Eng. Comput.*, vol. 58, no. 8, pp. 1767–1777, Aug. 2020.

[27] N. Wijethilake, M. Islam, D. Meedeniya, C. Chitraranjan, I. Perera, and H. Ren, "Radiogenomics of glioblastoma: Identification of radiomics associated with molecular subtypes," in *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-Oncology*. Lima, Peru: Springer, 2020, pp. 229–239.

[28] F. Tixier, H. Um, D. Bermudez, A. Iyer, A. Apte, M. S. Graham, K. S. Nevel, J. O. Deasy, R. J. Young, and H. Veeraraghavan, "Preoperative MRI-radiomics features improve prediction of survival in glioblastoma patients over MGMT methylation status alone," *Oncotarget*, vol. 10, no. 6, p. 660, 2019.

[29] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *NeuroImage*, vol. 31, no. 3, pp. 1116–1128, Jul. 2006.

[30] Y. Fu, N. Brown, S. Saeed, A. Casamitjana, Z. Baum, R. Delaunay, Q. Yang, A. Grimwood, Z. Min, S. Blumberg, J. Iglesias, D. Barratt, E. Bonmati, D. Alexander, M. Clarkson, T. Vercauteren, and Y. Hu, "DeepReg: A deep learning toolkit for medical image registration," *J. Open Source Softw.*, vol. 5, no. 55, p. 2705, Nov. 2020, doi: 10.21105/joss.02705.

[31] B. B. Avants, N. Tustison, and G. Song, "Advanced normalization tools (ANTS)," *Insight J*, vol. 2, no. 365, pp. 1–35, 2009.

[32] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012.

[33] B. Fischl, "FreeSurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, Aug. 2012.

[34] M. A. Hammoud, R. Sawaya, W. Shi, P. F. Thall, and N. E. Leeds, "Prognostic significance of preoperative MRI scans in glioblastoma multiforme," *J. Neuro-Oncol.*, vol. 27, no. 1, pp. 65–73, Jan. 1996.

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 234–241.

[36] I. D. Rubasinghe and D. A. Meedeniya, "Ultrasound nerve segmentation using deep probabilistic programming," *J. ICT Res. Appl.*, vol. 13, no. 3, pp. 241–256, 2019.

[37] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[38] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge," in *Proc. Int. MICCAI Brainlesion Workshop*. Quebec City, QC, Canada: Springer, 2017, pp. 287–297.

[39] M. Islam, V. S. Vibashan, V. J. M. Jose, N. Wijethilake, U. Utkarsh, and H. Ren, "Brain tumor segmentation and survival prediction using 3D attention UNet," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke Traumatic Brain Injuries*, A. Crimi and S. Bakas, Eds. Cham, Switzerland: Springer, 2020, pp. 262–272.

[40] L. Fidon, S. Ourselin, and T. Vercauteren, "Generalized wasserstein dice score, distributionally robust deep learning, and ranger for brain tumor segmentation: BraTS 2020 challenge," 2020, *arXiv:2011.01614*. [Online]. Available: http://arxiv.org/abs/2011.01614

[41] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *Proc. Int. MICCAI Brainlesion Workshop*. Granada, Spain: Springer, 2018, pp. 311–320.

[42] Z. Jiang, C. Ding, M. Liu, and D. Tao, "Two-stage cascaded U-net: 1st place solution to brats challenge 2019 segmentation task," in *Proc. Int. MICCAI Brainlesion Workshop*. Shenzhen, China: Springer, 2019, pp. 231–241.

[43] S. Saman and S. Jamjala Narayanan, "Survey on brain tumor segmentation and feature extraction of MR images," *Int. J. Multimedia Inf. Retr.*, vol. 8, no. 2, pp. 79–99, Jun. 2019.

[44] L. Sun, S. Zhang, H. Chen, and L. Luo, "Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning," *Frontiers Neurosci.*, vol. 13, no. 810, pp. 1–9, 2019.

[45] J. J. M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. H. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. W. L. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017.

[46] X. Feng, N. J. Tustison, S. H. Patel, and C. H. Meyer, "Brain tumor segmentation using an ensemble of 3D U-Nets and overall survival prediction using radiomic features," *Frontiers Comput. Neurosci.*, vol. 14, p. 25, Apr. 2020.

[47] 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," *Amer. J. Neuroradiol.*, vol. 41, no. 1, pp. 40–48, Jan. 2020.

[48] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[49] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," 2014, *arXiv:1405.3531*. [Online]. Available: http://arxiv.org/abs/1405.3531

[50] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *Proc. Int. MICCAI Brainlesion Workshop*. Quebec City, QC, Canada: Springer, 2017, pp. 178–190.

[51] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Athens, Greece: Springer, 2016, pp. 424–432.

[52] E. Puybareau, G. Tochon, J. Chazalon, and J. Fabrizio, "Segmentation of gliomas and prediction of patient overall survival: A simple and fast procedure," in *Proc. Int. MICCAI Brainlesion Workshop*. Granada, Spain: Springer, 2018, pp. 199–209.

[53] L. Chato, E. Chow, and S. Latifi, "Wavelet transform to improve accuracy of a prediction model for overall survival time of brain tumor patients based on MRI images," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Jun. 2018, pp. 441–442.

[54] Z. A. Shboul, M. Alam, L. Vidyaratne, L. Pei, M. I. Elbakary, and K. M. Iftekharuddin, "Feature-guided deep radiomics for glioblastoma patient survival prediction," *Frontiers Neurosci.*, vol. 13, p. 966, Sep. 2019.

[55] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," in *Proc. Annu. Conf. Med. Image Understand. Anal.* Edinburgh, U.K.: Springer, 2017, pp. 506–517.

[56] D. Nie, H. Zhang, E. Adeli, L. Liu, and D. Shen, "3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Athens, Greece: Springer, 2016, pp. 212–220.

[57] M. Incoronato, M. Aiello, T. Infante, C. Cavaliere, A. Grimaldi, P. Mirabelli, S. Monti, and M. Salvatore, "Radiogenomic analysis of oncological data: A technical survey," *Int. J. Mol. Sci.*, vol. 18, no. 4, p. 805, Apr. 2017.

[58] A. Welivita, I. Perera, D. Meedeniya, A. Wickramarachchi, and V. Mallawaarachchi, "Managing complex workflows in bioinformatics: An interactive toolkit with GPU acceleration," *IEEE Trans. Nanobiosci.*, vol. 17, no. 3, pp. 199–208, Jul. 2018.

[59] A. Alkuhlani, M. Nassef, and I. Farag, "Multistage feature selection approach for high-dimensional cancer data," *Soft Comput.*, vol. 21, no. 22, pp. 6895–6906, Nov. 2017.

[60] C. Hartmann, B. Hentschel, M. Tatagiba, J. Schramm, O. Schnell, C. Seidel, R. Stein, G. Reifenberger, T. Pietsch, A. von Deimling, M. Loeffler, and M. Weller, "Molecular markers in low-grade gliomas: Predictive or prognostic?" *Clin. Cancer Res.*, vol. 17, no. 13, pp. 4588–4599, Jul. 2011.

[61] S. Zuo, X. Zhang, and L. Wang, "A RNA sequencing-based six-gene signature for survival prediction in patients with glioblastoma," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, Dec. 2019.

[62] J. Xian, Q. Zhang, X. Guo, X. Liang, X. Liu, and Y. Feng, "A prognostic signature based on three non-coding RNAs for prediction of the overall survival of glioma patients," *FEBS Open Bio*, vol. 9, no. 4, pp. 682–692, 2019.

[63] W.-J. Zeng, Y.-L. Yang, Z.-Z. Liu, Z.-P. Wen, Y.-H. Chen, X.-L. Hu, Q. Cheng, J. Xiao, J. Zhao, and X.-P. Chen, "Integrative analysis of DNA methylation and gene expression identify a three-gene signature for predicting prognosis in lower-grade gliomas," *Cellular Physiol. Biochem.*, vol. 47, no. 1, pp. 428–439, 2018.

[64] P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, J. S. Barnholtz-Sloan, J. E. Velázquez Vega, D. J. Brat, and L. A. D. Cooper, "Predicting cancer outcomes from histology and genomics using convolutional networks," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 13, pp. E2970–E2979, Mar. 2018.

[65] S. Rathore, M. Aksam Iftikhar, and Z. Mourelatos, "Prediction of overall survival and molecular markers in gliomas via analysis of digital pathology images using deep learning," 2019, *arXiv:1909.09124*. [Online]. Available: http://arxiv.org/abs/1909.09124

[66] J. Zhao, L. Wang, D. Kong, G. Hu, and B. Wei, "Construction of novel DNA methylation-based prognostic model to predict survival in glioblastoma," *J. Comput. Biol.*, vol. 27, no. 5, pp. 718–728, May 2020.

[67] W.-Z. Gao, L.-M. Guo, T.-Q. Xu, Y.-H. Yin, and F. Jia, "Identification of a multidimensional transcriptome signature for survival prediction of postoperative glioblastoma multiforme patients," *J. Transl. Med.*, vol. 16, no. 1, p. 368, Dec. 2018.

[68] H. Gittleman, A. E. Sloan, and J. S. Barnholtz-Sloan, "An independently validated survival nomogram for lower-grade glioma," *Neuro-Oncology*, vol. 22, no. 5, pp. 665–674, May 2020.

[69] N. Beig, S. Singh, K. Bera, P. Prasanna, G. Singh, J. Chen, A. Saeed Bamashmos, A. Barnett, K. Hunter, V. Statsevych, V. B. Hill, V. Varadan, A. Madabhushi, M. S. Ahluwalia, and P. Tiwari, "Sexually dimorphic radiogenomic models identify distinct imaging and biological pathways that are prognostic of overall survival in glioblastoma," *Neuro-Oncology*, vol. 23, no. 2, pp. 251–263, Feb. 2021.

[70] R. J. Little and D. B. Rubin, *Statistical Analysis With Missing Data*, vol. 793. Hoboken, NJ, USA: Wiley, 2019.

[71] L. Ladha and T. Deepa, "Feature selection methods and algorithms," *Int. J. Comput. Sci. Eng.*, vol. 3, no. 5, pp. 1787–1797, 2011.

[72] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proc. 20th Int. Conf. Mach. Learn. (ICML)*, 2003, pp. 856–863.

[73] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 389–422, 2002.

[74] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2065, Apr. 2016, Art. no. 20150202.

[75] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[76] W. Rasanjana, S. Rajapaksa, I. Perera, and D. Meedeniya, "A SVM model for candidate Y-chromosome gene discovery in prostate cancer," in *11th Int. Conf. Bioinf. Comput. Biol. (BICOB)* 2019, pp. 129–138.

[77] K. E. Emblem, M. C. Pinho, F. G. Zöllner, P. Due-Tonnessen, J. K. Hald, L. R. Schad, T. R. Meling, O. Rapalino, and A. Bjornerud, "A generic support vector machine model for preoperative glioma survival associations," *Radiology*, vol. 275, no. 1, pp. 228–234, Apr. 2015.

[78] L. Macyszyn, H. Akbari, J. M. Pisapia, X. Da, M. Attiah, V. Pigrish, Y. Bi, S. Pal, R. V. Davuluri, L. Roccograndi, N. Dahmane, M. Martinez-Lage, G. Biros, R. L. Wolf, M. Bilello, D. M. O'Rourke, and C. Davatzikos, "Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques," *Neuro-Oncology*, vol. 18, no. 3, pp. 417–425, Mar. 2016.

[79] F. M. Khan and V. B. Zubek, "Support vector regression for censored data (SVRc): A novel tool for survival analysis," in *Proc. 8th IEEE Int. Conf. Data Mining*, Dec. 2008, pp. 863–868.

[80] E. Marubini, A. Morabito, and M. G. Valsecchi, "Prognostic factors and risk groups: Some results given by using an algorithm suitable for censored survival data," *Statist. Med.*, vol. 2, no. 2, pp. 295–303, Apr. 1983.

[81] A. Ciampi, R. Bush, M. Gospodarowicz, and J. 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*, vol. 47, no. 3, pp. 621–627, 1981.

[82] L. Gordon and R. A. Olshen, "Tree-structured survival analysis," *Cancer Treat. Rep.*, vol. 69, no. 10, pp. 1065–1069, 1985.

[83] A. Ciampi, J. Thiffault, J.-P. Nakache, and B. Asselain, "Stratification by stepwise regression, correspondence analysis and recursive partition: A comparison of three methods of analysis for survival data with covariates," *Comput. Statist. Data Anal.*, vol. 4, no. 3, pp. 185–204, Oct. 1986.

[84] A. Ciampi, "Recursive partition and amalgamation (RECPAM) for censored survival data: Criteria for tree selection," *Stat. Softw. Newslett.*, vol. 14, no. 2, pp. 78–81, 1988.

[85] W. J. Curran, C. B. Scott, J. Horton, J. S. Nelson, A. S. Weinstein, A. J. Fischbach, C. H. Chang, M. Rotman, S. O. Asbell, R. E. Krisch, and D. F. Nelson, "Recursive partitioning analysis of prognostic factors in three radiation therapy oncology group malignant glioma trials," *JNCI J. Nat. Cancer Inst.*, vol. 85, no. 9, pp. 704–710, May 1993.

[86] G. Bauman, K. Lote, D. Larson, L. Stalpers, C. Leighton, B. Fisher, W. Wara, D. Macdonald, L. Stitt, and J. G. Cairncross, "Pretreatment factors predict overall survival for patients with low-grade glioma: A recursive partitioning analysis," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 45, no. 4, pp. 923–929, Nov. 1999.

[87] H. Jun Cho and S.-M. Hong, "Median regression tree for analysis of censored survival data," *IEEE Trans. Syst., Man, Cybern., A, Syst. Humans*, vol. 38, no. 3, pp. 715–726, May 2008.

[88] M. L. Gandía-González, S. Cerdán, L. Barrios, P. López-Larrubia, P. G. Feijoó, A. Palpan, Jr., J. M. Roda, and J. Solivera, "Assessment of overall survival in glioma patients as predicted by metabolomic criteria," *Frontiers Oncol.*, vol. 9, p. 328, May 2019.

[89] J. R. Quinlan, "Bagging, boosting, and C4. 5," in *Proc. AAAI/IAAI*, vol. 1, 1996, pp. 725–730.

[90] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[91] B. Efron, "Tibshirani,(1993), an introduction to the bootstrap," in *Monographs on Statistics and Applied Probability*, vol. 1. Boca Raton, FL, USA: CRC Press, 1960.

[92] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006.

[93] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer, "Random survival forests," *Ann. Appl. Statist.*, vol. 2, no. 3, pp. 841–860, 2008.

[94] Y. S. Choi, S. S. Ahn, J. H. Chang, S.-G. Kang, E. H. Kim, S. H. Kim, R. Jain, and S.-K. and Lee, "Machine learning and radiomic phenotyping of lower grade gliomas: Improving survival prediction," *Eur. Radiol.*, vol. 30, pp. 3834–3842, Mar. 2020.

[95] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proc. Eur. Conf. Comput. Learn. Theory*. Barcelona, Spain: Springer, 1995, pp. 23–37.

[96] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[97] T. Hothorn, "Survival ensembles," *Biostatistics*, vol. 7, no. 3, pp. 355–373, Dec. 2005.

[98] J. Zhu, J. Chen, W. Hu, and B. Zhang, "Big learning with Bayesian methods," *Nat. Sci. Rev.*, vol. 4, no. 4, pp. 627–651, Jul. 2017.

[99] S. Rabinowicz, A. Hommersom, R. Butz, and M. Williams, "A prognostic model of glioblastoma multiforme using survival Bayesian networks," in *Proc. Conf. Artif. Intell. Med. Eur.* Vienna, Austria: Springer, 2017, pp. 81–85.

[100] M. Zhou, B. Chaudhury, L. O. Hall, D. B. Goldgof, R. J. Gillies, and R. A. Gatenby, "Identifying spatial imaging biomarkers of glioblastoma multiforme for survival group prediction," *J. Magn. Reson. Imag.*, vol. 46, no. 1, pp. 115–123, Jul. 2017.

[101] V. Bonato, V. Baladandayuthapani, B. M. Broom, E. P. Sulman, K. D. Aldape, and K.-A. Do, "Bayesian ensemble methods for survival prediction in gene expression data," *Bioinformatics*, vol. 27, no. 3, pp. 359–367, Feb. 2011.

[102] S. P. Luttrell, "A Bayesian analysis of self-organizing maps," *Neural Comput.*, vol. 6, no. 5, pp. 767–794, Sep. 1994.

[103] P. H. Abreu, M. S. Santos, M. H. Abreu, B. Andrade, and D. C. Silva, "Predicting breast cancer recurrence using machine learning techniques: A systematic review," *ACM Comput. Surv.*, vol. 49, no. 3, pp. 1–40, 2016.

[104] S.-M. Lee and P. A. Abbott, "Bayesian networks for knowledge discovery in large datasets: Basics for nurse researchers," *J. Biomed. Informat.*, vol. 36, nos. 4–5, pp. 389–399, Aug. 2003.

[105] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Mach. Learn.*, vol. 29, no. 2, pp. 131–163, Nov. 1997.

[106] M. Zhou, L. O. Hall, D. B. Goldgof, R. J. Gillies, and R. A. Gatenby, "Survival time prediction of patients with glioblastoma multiforme tumors using spatial distance measurement," *Proc. SPIE*, vol. 8670, Feb. 2013, Art. no. 86702O.

[107] S. R. Piccolo and L. J. Frey, "Clinical and molecular models of glioblastoma multiforme survival," *Int. J. Data Mining Bioinf.*, vol. 7, no. 3, p. 245, 2013.

[108] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, p. 386, 1958.

[109] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, Dec. 1943.

[110] D. Faraggi and R. Simon, "A neural network model for survival data," *Statist. Med.*, vol. 14, no. 1, pp. 73–82, Jan. 1995.

[111] T. Ching, X. Zhu, and L. X. Garmire, "Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data," *PLOS Comput. Biol.*, vol. 14, no. 4, Apr. 2018, Art. no. e1006076.

[112] S. Yousefi, F. Amrollahi, M. Amgad, C. Dong, J. E. Lewis, C. Song, D. A. Gutman, S. H. Halani, J. E. Velazquez Vega, D. J. Brat, and L. A. D. Cooper, "Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, Dec. 2017.

[113] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Amer. Stat. Assoc.*, vol. 53, no. 282, pp. 457–481, Jun. 1958.

[114] D. R. Cox, "Regression models and life-tables," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 34, no. 2, pp. 187–202, Jan. 1972.

[115] J. B.-K. Hsu, T.-H. Chang, G. A. Lee, T.-Y. Lee, and C.-Y. Chen, "Identification of potential biomarkers related to glioma survival by gene expression profile analysis," *BMC Med. Genomics*, vol. 11, no. S7, p. 34, Mar. 2019.

[116] L. Wang, Z. Yan, X. He, C. Zhang, H. Yu, and Q. Lu, "A 5-gene prognostic nomogram predicting survival probability of glioblastoma patients," *Brain Behav.*, vol. 9, no. 4, 2019, Art. no. e01258.

[117] T. Yang, P. Mao, X. Chen, X. Niu, G. Xu, X. Bai, and W. Xie, "Inflammatory biomarkers in prognostic analysis for patients with glioma and the establishment of a nomogram," *Oncol. Lett.*, vol. 17, no. 2, pp. 2516–2522, Dec. 2018.

[118] D. Koshland Jr, K. Walsh, and D. LaPorte, "Sensitivity of metabolic fluxes to covalent control," in *Current Topics in Cellular Regulation*, vol. 27. Amsterdam, The Netherlands: Elsevier, 1985, pp. 13–22.

[119] S. M. Lee, H.-J. Koh, D.-C. Park, B. J. Song, T.-L. Huh, and J.-W. Park, "Cytosolic $NADP^+$-dependent isocitrate dehydrogenase status modulates oxidative damage to cells," *Free Radical Biol. Med.*, vol. 32, no. 11, pp. 1185–1196, 2002.

[120] A. L. Cohen, S. L. Holmen, and H. Colman, "IDH1 and IDH2 mutations in gliomas," *Current Neurol. Neurosci. Rep.*, vol. 13, no. 5, pp. 765–773, May 2013.

[121] C. Houillier, X. Wang, G. Kaloshi, K. Mokhtari, R. Guillevin, J. Laffaire, S. Paris, B. Boisselier, A. Idbaih, F. Laigle-Donadey, K. Hoang-Xuan, M. Sanson, and J.-Y. Delattre, "IDH1 or IDH2 mutations predict longer survival and response to temozolomide in low-grade gliomas," *Neurology*, vol. 75, no. 17, pp. 1560–1566, Oct. 2010.

[122] B. Vogelstein, "P53 function and dysfunction," *Cell*, vol. 70, no. 4, pp. 523–526, Aug. 1992.

[123] N. Ishii, D. Maier, A. Merlo, M. Tada, Y. Sawamura, A.-C. Diserens, and E. G. Van Meir, "Frequent co-alterations of *TP*53, *p*16/*CDKN*2*A*, $p14^{ARF}$, PTEN tumor suppressor genes in human glioma cell lines," *Brain Pathol.*, vol. 9, no. 3, pp. 469–479, 1999.

[124] Y. Okamoto, P.-L. Di Patre, C. Burkhard, S. Horstmann, B. Jourde, M. Fahey, D. Schüler, N. M. Probst-Hensch, M. G. Yasargil, Y. Yonekawa, U. M. Lütolf, P. Kleihues, and H. Ohgaki, "Population-based study on incidence, survival rates, and genetic alterations of low-grade diffuse astrocytomas and oligodendrogliomas," *Acta Neuropathol.*, vol. 108, no. 1, pp. 49–56, Jul. 2004.

[125] X. Wang, J.-X. Chen, J.-P. Liu, C. You, Y.-H. Liu, and Q. Mao, "Gain of function of mutant TP53 in glioblastoma: Prognosis and response to temozolomide," *Ann. Surgical Oncol.*, vol. 21, no. 4, pp. 1337–1344, Apr. 2014.

[126] T. T. Batchelor, R. A. Betensky, J. M. Esposito, L.-D.-D. Pham, M. V. Dorfman, N. Piscatelli, S. Jhung, D. Rhee, and D. N. Louis, "Age-dependent prognostic effects of genetic alterations in glioblastoma," *Clin. Cancer Res.*, vol. 10, no. 1, pp. 228–233, Jan. 2004.

[127] K. Masutomi and W. C. Hahn, "Telomerase and tumorigenesis," *Cancer Lett.*, vol. 194, no. 2, pp. 163–172, May 2003.

[128] M. Labussière, A. L. Di Stefano, V. Gleize, B. Boisselier, M. Giry, S. Mangesius, A. Bruno, R. Paterra, Y. Marie, A. Rahimian, G. Finocchiaro, R. S. Houlston, K. Hoang-Xuan, A. Idbaih, J.-Y. Delattre, K. Mokhtari, and M. Sanson, "TERT promoter mutations in gliomas, genetic associations and clinico-pathological correlations," *Brit. J. Cancer*, vol. 111, no. 10, pp. 2024–2032, Nov. 2014.

[129] B. Heidenreich, P. S. Rachakonda, K. Hemminki, and R. Kumar, "TERT promoter mutations in cancer development," *Current Opinion Genet. Develop.*, vol. 24, pp. 30–37, Feb. 2014.

[130] P. J. Killela, Z. J. Reitman, and Y. Jiao, "TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 15, pp. 6021–6026, Apr. 2013.

[131] H. Arita, Y. Narita, S. Fukushima, K. Tateishi, Y. Matsushita, A. Yoshida, Y. Miyakita, M. Ohno, V. P. Collins, N. Kawahara, S. Shibui, and K. Ichimura, "Upregulating mutations in the TERT promoter commonly occur in adult malignant gliomas and are strongly associated with total 1p19q loss," *Acta Neuropathol.*, vol. 126, no. 2, pp. 267–276, Aug. 2013.

[132] N. Nonoguchi, T. Ohta, J.-E. Oh, Y.-H. Kim, P. Kleihues, and H. Ohgaki, "TERT promoter mutations in primary and secondary glioblastomas," *Acta Neuropathol.*, vol. 126, no. 6, pp. 931–937, Dec. 2013.

[133] Y. Lee, J. Koh, S.-I. Kim, J. K. Won, C.-K. Park, S. H. Choi, and S.-H. Park, "The frequency and prognostic effect of TERT promoter mutation in diffuse gliomas," *Acta Neuropathol. Commun.*, vol. 5, no. 1, p. 62, Dec. 2017.

[134] J. Liu, X. Zhang, X. Yan, M. Sun, Y. Fan, and Y. Huang, "Significance of TERT and ATRX mutations in glioma," *Oncol. Lett.*, vol. 17, no. 1, pp. 95–102, Oct. 2018.

[135] K. Gao, G. Li, Y. Qu, M. Wang, B. Cui, M. Ji, B. Shi, and P. Hou, "TERT promoter mutations and long telomere length predict poor survival and radiotherapy resistance in gliomas," *Oncotarget*, vol. 7, no. 8, p. 8712, 2016.

[136] A. E. Pegg, M. E. Dolan, and R. C. Moschel, "Structure, function, and inhibition of O6-alkylguanine-DNA alkyltransferase," in *Progress in Nucleic Acid Research and Molecular Biology*, vol. 51. Amsterdam, The Netherlands: Elsevier, 1995, pp. 167–223.

[137] S. C. Schold, D. M. Kokkinakis, J. L. Rudy, R. C. Moschel, and A. E. Pegg, "Treatment of human brain tumor xenografts with O6-benzyl-2'-deoxyguanosine and BCNU," *Cancer Res.*, vol. 56, no. 9, pp. 2076–2081, 1996.

[138] W. Wick, M. Weller, M. Van Den Bent, M. Sanson, M. Weiler, A. Von Deimling, C. Plass, M. Hegi, M. Platten, and G. Reifenberger, "*MGMT* testing—The challenges for biomarker-based glioma treatment," *Nature Rev. Neurol.*, vol. 10, no. 7, p. 372, 2014.

[139] R. G. W. Verhaak, K. A. Hoadley, E. Purdom, V. Wang, and Y. Qi, "Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1," *Cancer Cell*, vol. 17, no. 1, pp. 98–110, Jan. 2010.

[140] H. S. R. Rajula, G. Verlato, M. Manchia, N. Antonucci, and V. Fanos, "Comparison of conventional statistical methods with machine learning in medicine: Diagnosis, drug development, and treatment," *Medicina*, vol. 56, no. 9, p. 455, Sep. 2020.

[141] S. Zhao, W.-P. Fung-Leung, A. Bittner, K. Ngo, and X. Liu, "Comparison of RNA-seq and microarray in transcriptome profiling of activated T cells," *PLoS ONE*, vol. 9, no. 1, Jan. 2014, Art. no. e78644.

[142] Z. Bodalal, S. Trebeschi, T. D. L. Nguyen-Kim, W. Schats, and R. Beets-Tan, "Radiogenomics: Bridging imaging and genomics," *Abdominal Radiol.*, vol. 44, no. 6, pp. 1960–1984, Jun. 2019.

[143] F. Q. Wiki and Q. F.-P. W. Body, "The radiological society of North America (RSNA) quantitative imaging biomarker alliance (QIBA) FDG-PET technical committee: Proffered protocol for quantifying the standard uptake (SUV) in patients with cancer," no. Version 1.13, Tech. Rep., 2016.

[144] M. M. Li, M. Datto, E. J. Duncavage, S. Kulkarni, N. I. Lindeman, S. Roy, A. M. Tsimberidou, and C. L. Vnencak-Jones, "Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A joint consensus recommendation of the association for molecular pathology, American Society of clinical Oncology, and College of American pathologists," *J. Mol. Diagnostics*, vol. 19, no. 1, pp. 4–23, 2017.

[145] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[146] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," 2013, *arXiv:1312.6034*. [Online]. Available: http://arxiv.org/abs/1312.6034

[147] A. Welivita, I. Perera, and D. Meedeniya, "An interactive workflow generator to support bioinformatics analysis through GPU acceleration," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 457–462.

[148] S. Rajapaksa, W. Rasanjana, I. Perera, and D. Meedeniya, "GPU accelerated maximum likelihood analysis for phylogenetic inference," in *Proc. 8th Int. Conf. Softw. Comput. Appl.*, New York, NY, USA, Feb. 2019, pp. 6–10.

[149] T. Karn, "High-throughput gene expression and mutation profiling: Current methods and future perspectives," *Breast Care*, vol. 8, no. 6, pp. 401–406, 2013.

[150] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Granada, Spain: Springer, 2018, pp. 92–104.

[151] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen, and S. Bakas, "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Dec. 2020.

NAVODINI WIJETHILAKE (Student Member, IEEE) received the B.Sc. degree in engineering from the University of Moratuwa, Sri Lanka, where she is currently pursuing the master's degree. She has worked as a Research Intern with the Medical Mechatronics Laboratory, National University of Singapore.

DULANI MEEDENIYA (Member, IEEE) received the Ph.D. degree in computer science from the University of St Andrews, U.K. She is currently a Senior Lecturer with the Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka. Her main research interests include software modeling and design, bio-health informatics, machine learning, data analytics and technology enhanced learning. She is a Fellow of HEA(UK), MIET, MIEEE, and a Charted Engineer registered at EC (UK).

CHARITH CHITRARANJAN received the B.Sc. Eng. degree (Hons.) from the University of Moratuwa, Sri Lanka, and the M.Sc. and Ph.D. degrees from NDSU. He is currently a Lecturer with the University of Moratuwa. His research interests include data science biological, motor traffic-related data, and MNBD.

INDIKA PERERA (Senior Member, IEEE) received the B.Sc.Eng. (Hons.) and M.Sc. degrees from the University of Moratuwa, Sri Lanka, and the M.B.S. and P.G.D.B.M. degrees from the University of Colombo, and the Ph.D. degree from the University of St Andrews, U.K. He is currently a Senior Lecturer with the University of Moratuwa. His research interests include research topics of software architecture, software engineering; technology enhanced learning, UX and application development for bio-health research. He is a Fellow of HEA, U.K., MIET, SMIEEE, and a Charted Engineer registered at EC, U.K., and IE, Sri Lanka.

MOBARAKOL ISLAM (Member, IEEE) received the Ph.D. degree from the National University of Singapore, Singapore, in 2019. He is currently working with the Biomedical Image Analysis Group, Imperial College London, U.K. His current research interests include causal DL, medical imaging, and image-guided surgery.

HONGLIANG REN (Senior Member, IEEE) has navigated his academic journey through The Chinese University of Hong Kong, Johns Hopkins University, Children's Hospital Boston, Harvard Medical School, Children's National Medical Center, USA, and the National University of Singapore. His research interests include biorobotics, intelligent control, medical mechatronics, soft continuum robots, soft sensors, and multisensory learning in medical robotics. He was a recipient of the NUS Young Investigator Award and Engineering Young Researcher Award, IAMBE Early Career Award 2018, Interstellar Early Career Investigator Award 2018, and ICBHI Young Investigator Award 2019. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATION SCIENCE & ENGINEERING (T-ASE) and *Medical & Biological Engineering & Computing* (MBEC).