Predicting Genetic Markers for Brain Tumors Using a Composite Loss

Arijit De, *Student Member, IEEE*, Aritro Santra, Mona Tiwari, and Ananda S. Chowdhury, *Senior Member, IEEE*

***Abstract*—Brain cancer has a very high mortality rate. Gliomas are the most common malignant brain tumors, causing this severe fatality. Recent biological investigations revealed that a holistic study of biomarkers, responsible for causing genetic mutations in gliomas, can ensure a comprehensive prognosis and treatment plan for the patients. In this paper, we simultaneously predict five such important genetic markers, namely, IDH, 1p/19q codeletion status, ATRX, MGMT, and TERT from Whole Slide Images using deep learning. At the heart of our deep learning based solution lies a novel composite loss function, which uniquely combines individual, pairwise and groupwise traits of the above bio-markers. While multi-label weighted cross-entropy loss captures individual characteristics, a conditional probability loss was developed for pairwise behavior, and, a spectral graph loss was formulated for modeling group properties. Comprehensive experimentation, along with an ablation study, clearly demonstrates the effectiveness of our solution, by achieving state-of-the-art prediction.**

***Index Terms*—Genetic markers, prediction, brain tumor, whole slide image, composite loss.**

I. Introduction

BRAIN tumors account for a disproportionate burden of cancer mortality because of their high fatality rate; only one-third of individuals survive at least 5 years after diagnosis [1]. Gliomas are the most common primary malignant brain tumors in adults, arising from glial cells. More recently, several molecular signatures have been identified to predict favorable patient response to treatment as well as overall outcome in gliomas [2]. Collective study of various genetic mutations in gliomas, which are captured through different biomarkers, are quite essential for a comprehensive prognosis and subsequent treatment plan [3].

Some works are reported in the past on identification of biomarkers from radiological data, like MRI, using machine learning techniques [4], [5]. But there are inherent limitations to using radiological data for predicting genetic markers. Firstly, genetic markers are traditionally identified by examining histopathological data, like Whole Slide Image (WSI) samples, as the effect of genetic mutations are more visible at the cellular level [6]. Secondly, as mentioned in [7], though a model can be trained to learn the correlation between radiological features and the biomarker status; WSI data can offer morphological features, that help a model learn better the distinguishing factors of the biomarkers. WSI offers a better explainability for models trained to identify glioma biomarkers [8]. As shown in Fig. 1, there are three different types of glioma tissues which exhibit quite different visual characteristics. Each of them has a different combination of biomarkers. A model can typically learn these combinations by capturing suitable features from WSI. Some related problems have been quite well investigated. For example, multi-class brain tumor classification is studied in recent times in [9]. Multi-class biomarker detection [10] and prediction [11] using WSI image have also been investigated. In such problems, one typically needs to focus on a single genetic marker, which is solely responsible for tumor mutation. However, a more challenging, and biologically more intriguing problem, is to predict multi-label genomic biomarker, where multiple genetic markers collectively influence tumor mutation. Naturally, the complexity of the problem increases with an increase in the number of biomarkers to be detected. This class of problem remains largely unexplored. Le et. al. [12] predicted 9 such biomarkers, but they did not consider the inter-biomarker relationships.

We now discuss the biological importance of the five glioma genetic markers, which are predicted in this work. These markers are Isocitrate DeHydrogenase (**IDH**), combined loss of the short arm of chromosome 1 and the long arm of chromosome 19 (**1p/19q codeletion**), O(6)- methylguanine-DNA methyltransferase (**MGMT**), Alpha- Thalassemia/mental Retardation, X-linked (**ATRX**), and, Telomerase Reverse TranscripTase (**TERT**) in the context of gliomas. Literature review suggests that these biomarkers can affect glioma *individually*, in a *pairwise* manner, and, as a *group*. IDH genes play a pivotal role in low grade gliomas (LGG) with favorable prognosis [13]. 1p/19q codeletion leads to better prognosis in oligodendroglial tumors [14]. MGMT is a vital DNA repair protein. In glioblastoma patients, MGMT promoter methylation status is a significant marker for therapeutic response to temozolomide, an alkylating agent [15]. Furthermore, ATRX mutations are also important as they are associated with increased survival in

Received 18 June 2024; revised 15 July 2025; accepted 19 July 2025. Date of publication 28 July 2025; date of current version 10 December 2025. The work of Arijit De was supported by the Research Scholar Program of Tata Consultancy Services (TCS-RSP). *(Corresponding author: Ananda S. Chowdhury.)*

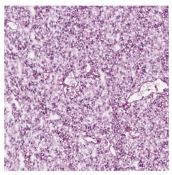
Arijit De is with Electronics and Telecommunication Engineering Department, Jadavpur University, Kolkata 700032, India (e-mail: arijitde.etce.rs@jadavpuruniversity.in).

Aritro Santra is with the Electronics and Communication Engineering Department, Manipal Institute of Technology, Manipal 576104, India (e-mail: santra.aritro@gmail.com).

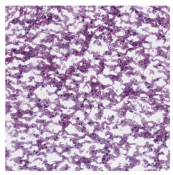
Mona Tiwari is with the Department of Neuro-Radiology, Institute of Neurosciences, Kolkata 700017, India (e-mail: drmonatiwari@gmail.com).

Ananda S. Chowdhury is with Electronics and Telecommunication Engineering Department, Jadavpur University, Kolkata 700032, India (e-mail: as.chowdhury@jadavpuruniversity.in).

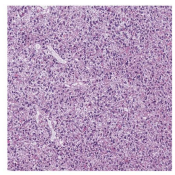
Digital Object Identifier 10.1109/TCBBIO.2025.3593318



(a) IDH-wildtype, No 1p/19q codeletion, ATRX-wildtype, TERT mutated and MGMT methylated



(b) IDH-mutant, No 1p/19q codeletion, ATRX-mutant, TERT unmutated and MGMT methylated



(c) IDH-mutant, 1p/19q codeletion, ATRX-wildtype, TERT mutated and MGMT methylated

Fig. 1. Three glioma tumor tissue samples having different combinations of biomarkers.

individuals with astrocytic tumors [16]. Finally, in the genomic landscape of gliomas, mutations in the TERT promoter region have emerged as a common occurrence [17].

As a part of pairwise behavior, one biomarker influences the presence or absence of another biomarker in case of gliomal tumors. For example, as mentioned in [18], 1p/19q codeletions are closely linked to IDH mutation. As demonstration of a composite behavior, multiple biomarkers are found to influence different types of glioma. For example, the presence of a TERT mutation without IDH mutation and 1p/19q co-deletion suggests the occurence of Glioblastoma [19]. Such composite behaviors are also captured by three glioma tumor tissue samples in Fig. 1.

In this work, we simultaneously predict five different glioma causing biomarkers, i.e., IDH, 1p/19q codeletion status, ATRX, MGMT and TERT from WSI images using a deep network. A comprehensive knowledge of these five biomarkers can ensure faster and better prognosis and treatment plan for the patients. In our solution, we explicitly capture the individual, pairwise, and, group behavior of the biomarkers as revealed by the existing literature. A composite loss function is designed by combining these three losses. We did not find any public dataset that has directly made available a mapping of WSI image and the corresponding biomarker information. This mapping is crucial in order to learn label correlation information from bio-markers along with image features from WSI. We carefully analyze, clean and select data samples that had both biomarker and WSI sample for the same patient, and, create a new benchmark dataset for ready use. Our contributions are now summarized below:

- We simultaneously predict five glioma causing biomarkers. In our deep learning based solution model, we explicitly capture their individual traits as well as pairwise and collective interrelations with a composite loss function. The three components of this loss function are as follows:
  - A multi-label weighted cross-entropy loss, with weights for each bio-marker based on their probability of occurrence. We create a five-headed FCN which fine-tunes the model to calculate the multi-label weighted cross-entropy loss for each of the five individual biomarkers. This teaches the model to predict bio-markers based on their *individual traits*.
  - A conditional probability based loss function which harnesses the cause-effect relationship of the biomarker pairs. This teaches the model to predict biomarkers based on their *pair-wise relationship*.
  - A graph similarity based loss function that makes the model learn the bio-marker co-occurrence information. This teaches the model to predict biomarkers based on the *composite (group-wise) relation* among them.
- We build a common benchmark data repository with WSI image and the corresponding biomarker information by integrating WSI data from The Cancer Imaging Archive (TCIA) [20], [21] with biomarker data from the Genomic data commons (GDC) [22] of the same patient. We believe this type of benchmark will be beneficial for studying multiple biomarker predictions of glioma. This dataset can be publicly accessed using this link.
- We demonstrate through comprehensive experiments that we achieve state-of-the-art prediction results by outperforming a number of competing methods on the benchmark dataset.

II. Related Work

We have come across some works, which have used machine learning techniques for the prediction of genomic mutations. However, most of these studies are restricted to separately predicting specific mutations only.

In this context, we first discuss the works that have used MRI data for the prediction of genetic biomarkers in gliomas. To predict the MGMT status, the authors in [23] used multi-modal MRIs, and various deep learning architectures. However, the reported results were not to the mark. In [24], the authors employed various data preprocessing techniques, and applied deep learning models to find the MGMT status. The reported results were again below par. In another work [25], radiogenomic features were extracted from MRI data to classify the 1p/19q status. Some works have attempted to predict two biomarkers from MRI data. Zhou et al. [26] applied a random forest approach to initially determine IDH status. They then trained a second model to predict 1p/19q status within the IDH mutant subset. Nishikawa et al. [27] used 3D deep learning models on MRI data to simultaneously predict the status of IDH and 1p/19q codeletion status. Tripathi and Bag [5] performed classification of tumor grade along with the determination of IDH and 1p/19q status using MRI data. But MRI data provides information and features about the tumor macro environment, which a model uses to learn the correlation between label and data [7]. Further, the tumor region needs to be segmented to learn the correlation, which requires additional processing and hence increases the computational burden. Still, the model does not really learn the features of cellular structures, which are directly changed as a result of biomarker mutation, as shown in [6]. Thus, it fails to provide valuable information about other biomarkers

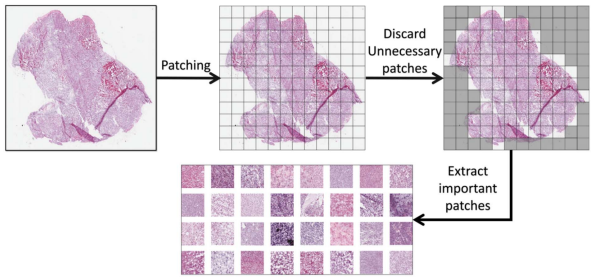


Fig. 2. Patch generation process: A whole slide image is divided into equal sized square patches. Each patch is scanned and the uninformative patches (shown in gray) are discarded. Remaining patches containing actual tissue images are saved.

like MGMT, ATRX and TERT which are analysed with WSI data in real clinical scenario. This fact is corroborated by [23], where, they did not achieve desirable prediction results for MGMT status from MRI. Similarly, as discussed in [28], the morphological effects of ATRX and TERT mutations are clearly visible in WSI images.

We also explored some works that have used WSI as their imaging modality. For IDH status prediction, Liechty et al. [29] used uniform manifold approximation and projection (UMAP) on patch-level embedding vectors of WSI images and achieved decent results. For the same task, Liu et al. [30] used multiscale feature fusion and arrived at even better results. In [13], the authors used GAN and incorporated patient specific information like age to predict the presence of IDH mutations. Wang et al. [31], through their experiments using both MRI and WSI data, showed that WSI data alone provides better classification accuracy for IDH status prediction than that of from MRI. All these works clearly endorse the fact that WSI is certainly a better modality for prediction of genetic markers (as compared to MRI). The above review also indicates that there is a marked paucity of published works on simultaneous prediction of multiple mutations, which is very much essential to plan fast and efficient prognosis. Le et al. [12] predicted nine biomarkers responsible for glioma. While this is certainly an important work, the authors did not capture the complex inter-relationships (pair-wise, as well as group-wise) among different biomarkers.

Note that the task of predicting multiple biomarkers is essentially a multi-label prediction problem. We now discuss some loss functions, which are applied for multi-label prediction tasks. Wu et al. [32] considered underlying distribution of data and class imbalance to improve the performance of their model. However, they used a binary cross-entropy based loss function. Further, their technique is more suitable for datasets having long-tailed distributions, which is not the case here. Ridnik et al. [33] employed asymmetric loss to reduce the negative effect of class imbalance. They used Regions of Interest (ROI), whereas our model learns from the whole image. Yessou et al. [34] compared various multi-label loss functions. However, none of these functions have incorporated any label correlation knowledge, a very important factor in medical image analysis [35]. Finally, the authors in [12] employed only cross-entropy loss.

As a summary, we can say that our work, which focuses on simultaneously predicting five crucial genetic biomarkers for brain tumors from WSI data, captures individual, pair-wise, and group-wise behavior (of these biomarkers). Accordingly, we design a composite loss function, with cross-entropy, conditional probability-based, and graph spectral loss articulating different behaviors of these markers. Through comprehensive experiments, we demonstrate that our method yields superior results as compared to the state-of-the-art methods.

III. Methodology

In this section, we first describe how the image is fed to the model, followed by the design of loss functions. We then discuss how we conduct the training and testing of the model.

A. Image Feature Extraction

We use a CNN to extract features from WSI. Since, a single color WSI has extremely high resolution (e.g., $\sim 20000 \times 20000 \times 3$), it becomes impractical to directly feed the image into a CNN. Hence, we divide the image into a bag of smaller equal sized patches, following [36]. Another reason to use multiple different patches from the same WSI image is to increase the sample size for training. Since, WSI is captured from the actual tissue slide, it contains several unnecessary artifacts like bubbles, dust, and, cracks. Furthermore, as the background (region without any tissue image) is white, the extracted smaller white patches are of no use. So, we discarded all the unnecessary patches using the same strategy as adopted in [37]. The patch generation process is shown in Fig. 2. WSI at different magnifications convey different feature information. To utilize these different types of information, patches are extracted from $20\times$ (local level) and $5\times$ (global level) magnification, similar to [38]. Lower magnification levels preserve overall morphological structures, while higher magnification levels capture finer details [39]. Let each such input image patch of $20\times$ magnification of a sample WSI be denoted by I_i^l $(i = 1, \ldots, n_l)$, where n_l is the total number of local image patches in the bag. Similarly, each image patch of $5\times$ magnification level of the same WSI is denoted by I_i^g $(i = 1, \ldots, n_g)$, where n_g is the total number of global image patches in the bag. Suppose, F_i^l and F_i^g be d-dimensional feature vectors for I_i^l and, I_i^g respectively, obtained using a CNN.

The F_i^l's and F_i^g's are next inputted to each FCN. All the FCNs have a sigmoid layer in the end. Let the b-dimensional probability vectors denoting the occurrence of the biomarkers for I_i^l, and, I_i^g be respectively represented by p_i^l and, p_i^g. For this problem, $b = 5$. The p_i^l's and p_i^g's vectors are obtained from the b number of FCNs. We further construct binary prediction vectors B_i^l and B_i^g from p_i^l and p_i^g. Please see Table I for the biomarkers and the interpretation of the values of 1 and 0. For the sake of brevity, we will henceforth simply use B to denote a prediction vector. All the constituent image patches of any particular WSI sample have the same annotation (ground-truth). Hence, they have the same reference vector, say, B_r. In this way, we have the training patches, with a prediction vector, and, a reference prediction vector for all.

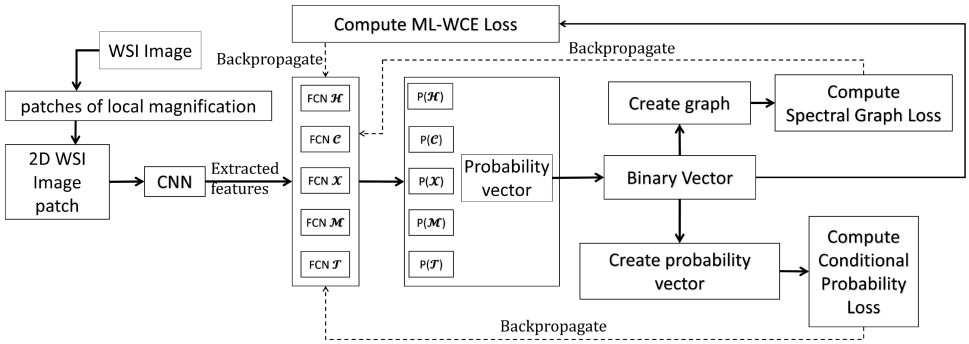


Fig. 3. Training workflow for a single WSI image at local magnification level. This is repeated for other WSI images in the training set at local magnification level. The whole training process is then repeated at global magnification level.

TABLE I
DETAILS OF THE GENETIC MARKERS: ID, NAME, AND, INTERPRETATION OF THE VALUES OF 1 AND 0 FOLLOWING [19]

ID	Name	Value: 1	Value: 0
$\mathcal{H}$	IDH	Mutant	Wild-type
$\mathcal{C}$	1p/19q codeletion	Codeleted	Non-codeleted
$\mathcal{X}$	ATRX	Mutant	Wild-type
$\mathcal{M}$	MGMT	Methylated	Unmethylated
$\mathcal{T}$	TERT	Mutant	Wild-type

B. Design of Loss Function

We design three different loss functions to capture individual, pairwise and groupwise traits of the biomarkers. These three losses are eventually combined to form a composite loss.

1) Individual Multi-Label Weighted Cross-Entropy Loss: Multi-label Weighted Cross-Entropy (ML-WCE) loss assigns weights to the bio-markers based on the number of samples in the dataset for which they are present and absent. This captures impacts of the bio-markers, solely at individual level. In our dataset, the bio-markers are not evenly distributed among the samples. As this is a multi-label problem, we need to obtain separate weights for different bio-markers. Let the number of positive samples for the bio-marker i be s_i and the number of negative samples for the same be t_i. A positive sample means a value 1 for the sample, and, a negative sample means a value 0. Please see Table I for the interpretation. The probability of presence of the bio-marker i can be calculated as $w_i^s = \frac{s_i}{m}$, and, the probability of absence of the bio-marker i as $w_i^t = \frac{t_i}{m}$, where m is the total number of WSI samples. Standard cross-entropy loss is given by:

$$\mathcal{L}_C = -\left[y_i \times \log(\hat{y}_i) + (1 - y_i) \times \log(1 - \hat{y}_i)\right] \quad (1)$$

where, y_i and $\hat{y}_i$ are the actual and predicted labels for the biomarker i respectively [40]. In this problem, we modify the above equation by adding the two weights w_i^s and w_i^t. Let us denote the modified cross-entropy loss for the bio-marker i by Ψ_i. So, we write:

$$\Psi_i = -\left[w_i^s \times y_i \times log(\hat{y}_i) + w_i^t \times (1 - y_i) \times log(1 - \hat{y}_i)\right] \quad (2)$$

Each individual Ψ_i is calculated for each individual FCN in order to learn a specific biomarker as shown in Fig. 3. The loss for all the bio-markers is obtained as the summation over all Ψ_i. So, we write:

$$\mathcal{L}_{\mathcal{C}} = \sum_{i=1}^{b} \Psi_i \quad (3)$$

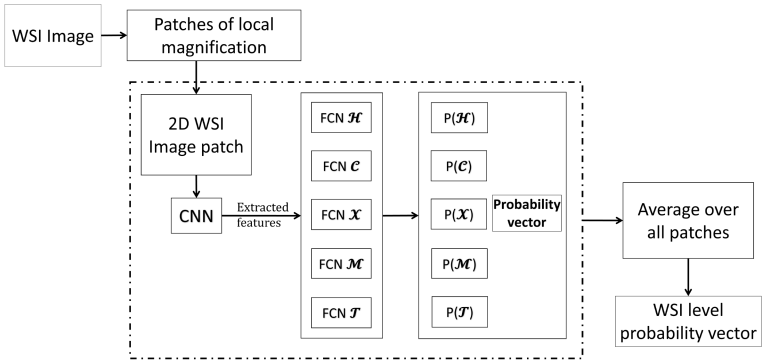
The training process aims to minimize $\mathcal{L}_C$ so that B_r and B becomes more similar as training progresses, with the ideal goal of making B equal to B_r.

2) Pairwise Conditional Probability Loss: We next formulate a conditional probability based loss function to model pairwise behavior of the biomarkers. The conditional probability links the presence/absence of one bio-marker caused by presence/absence of a second bio-marker.

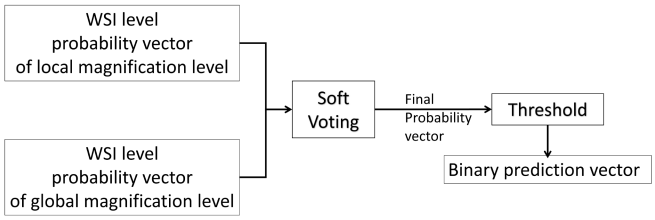
Let the estimated conditional probability be denoted by, $\mathcal{P}(i = \tau_1 | j = \psi_1)$ where i, j are the bio-markers ($i = 1, \ldots, b; j = 1, \ldots, b; i \neq j$). Let the actual conditional probability, which is given as ground-truth, be denoted by $\mathcal{P}_r(i = \tau_2 | j = \psi_2)$. Here, $\tau_1, \tau_2, \psi_1, \psi_2 \in \{0, 1\}$. We show in the Appendix with an example how $\mathcal{P}$ can be obtained from B. Likewise, one can obtain $\mathcal{P}_r$ from B_r. The conditional probability loss is deemed as the absolute difference between the sum of estimated conditional probabilities and actual conditional probabilities and, can be expressed as:

$$\mathcal{L}_{\mathcal{P}} = \sum_{i=1}^{b} \sum_{j=1}^{b} \left(|\mathcal{P}_r(i = \tau_1 | j = \psi_1) - \mathcal{P}(i = \tau_2 | j = \psi_2)| \times \delta_{\tau_1, \tau_2} \times \delta_{\psi_1, \psi_2}\right) \quad (4)$$

In the above equation, δ_{τ_1, τ_2} is equal to 1 if τ_1 and τ_2 are equal, and is 0 otherwise. Similar is the interpretation for δ_{ψ_1, ψ_2}. For the present problem, we selected 5 sets of conditional probabilities, which are rendered biologically important by the experts. As mentioned in [41], TERT mutations commonly occurs with 1p/19q codeletion, therefore the probability $\mathcal{P}(\mathcal{T} = 1 | \mathcal{C} = 1)$ is high and is an important relation to consider. Similarly, as mentioned in [18], 1p/19q co-deletions are mutually exclusive with ATRX mutations, i.e., $\mathcal{P}(\mathcal{C} = 1 | \mathcal{X} = 0)$ and $\mathcal{P}(\mathcal{C} = 0 | \mathcal{X} = 1)$ are important relations. We likewise formulate the rest of the pairwise relations based on evidences given in [18], [19]. Hence,



(a) Testing workflow showing how we get the WSI level probability vector by averaging the probability vectors of all images for the sample under consideration. This is repeated at both local and global magnification levels.



(b) Combining WSI level probability vectors of local and global magnification levels into a single binary prediction vector.

Fig. 4. Overall test time workflow of our model. (a) shows how we get a single probability vector for a single WSI image for local magnification level. This process is repeated at the global magnification level also. (b) shows how we derive the final binary prediction from combining the results of (a).

for this work, the general equation in (4) can be simplified to:

$$\begin{aligned}\mathcal{L}_{\mathcal{P}} = \; & |\mathcal{P}_r(\mathcal{T}=1|\mathcal{H}=1) - \mathcal{P}(\mathcal{T}=1|\mathcal{H}=1) \\ & + \mathcal{P}_r(\mathcal{M}=1|\mathcal{H}=1) - \mathcal{P}(\mathcal{M}=1|\mathcal{H}=1) \\ & + \mathcal{P}_r(\mathcal{C}=1|\mathcal{X}=0) - \mathcal{P}(\mathcal{C}=1|\mathcal{X}=0) \\ & + \mathcal{P}_r(\mathcal{C}=0|\mathcal{X}=1) - \mathcal{P}(\mathcal{C}=0|\mathcal{X}=1) \\ & + \mathcal{P}_r(\mathcal{T}=1|\mathcal{C}=1) - \mathcal{P}(\mathcal{T}=1|\mathcal{C}=1) \\ & + \mathcal{P}_r(\mathcal{X}=1|\mathcal{H}=0) - \mathcal{P}(\mathcal{X}=1|\mathcal{H}=0)| \end{aligned} \tag{5}$$

where, $\mathcal{H}, \mathcal{C}, \mathcal{X}, \mathcal{M}$ and $\mathcal{T}$ refer to the bio-markers as shown in Table I. The training process aims to minimize $\mathcal{L}_P$.

Now, we show how to calculate the probability $\mathcal{P}(i=\tau_1|j=\psi_1)$ for any two biomarkers τ and ψ. To obtain $\mathcal{P}(i=\tau_1|j=\psi_1)$, we need the predicted binary vector B. Since, we get a batch of Bs as output during training, we can calculate $\mathcal{P}(i=\tau_1)$, $\mathcal{P}(j=\psi_1)$ and $\mathcal{P}(i=\tau_1 \cap j=\psi_1)$ by first finding the number of times $i=\tau_1$ occurs, $j=\psi_1$ occurs, and, $i=\tau_1$ and $j=\psi_1$ jointly occur, respectively; and then dividing each by the batch size. An example of how to find $\mathcal{P}(i=1|j=0)$ is presented below.

Let us consider a batch with size $z_1+z_2+z_3+z_4$. Here, z_1 is the number of occurrences where both i and j have values 1; z_2 is the number of occurrences where i has a value 1 but j has a value 0; z_3 denotes the number of cases where i has a value 0 but j has a value 1; and, z_4 is the number of cases where both i and j have values 1.

$$\mathcal{P}(j=0) = \frac{z_1+z_2}{z_1+z_2+z_3+z_4} \tag{6}$$

$$\mathcal{P}(i=1 \cap j=0) = \frac{z_2}{z_1+z_2+z_3+z_4} \tag{7}$$

Hence,

$$\mathcal{P}(i=1|j=0) = \frac{\mathcal{P}(i=1 \cap j=0)}{\mathcal{P}(j=0)} = \frac{z_2}{z_1+z_2} \tag{8}$$

3) Group-Wise Spectral Graph Loss: We introduce here a third type of loss using a graph-based formulation. This loss aims to make the model learn about co-occurrence of multiple bio-markers. The group information of the five genetic markers is captured through a weighted complete graph $G=G(V,E)$. Each vertex of this graph denotes a genetic marker as explained in Table I. So, $V=[v_1, v_2, \ldots, v_b]$ and $|V|=b$. We previously obtain the b-dimensional binary prediction vector, B. This means, $B[i]$ denotes the value of v_i and likewise. In the complete graph, we represent the affinity of the bio-markers by assigning appropriate weights to the edges. For example, in a particular sample, if three out of the five biomarkers occur together, then the edges connecting them are assigned high values. In contrast, the edges connecting these vertices with the two remaining vertices are given low values. This is shown in Fig. 5, where, three different graphs are built from the genetic information of the sample images shown in Fig. 1. We set the edge weights (w_{ij}) between the vertices v_i and v_j using the following equation:

$$w_{i,j} = e^{-2|v_i - v_j|} \tag{9}$$

So, $|v_i - v_j|$ can assume a value of either 0 (when v_i and v_j have same values) or 1 (when v_i and v_j have opposite values). Exponentiation is introduced to provide more discrimination between these two cases, making the edge weights either 1 (when v_i and v_j have same values) or e^{-2} (when v_i and v_j have opposite

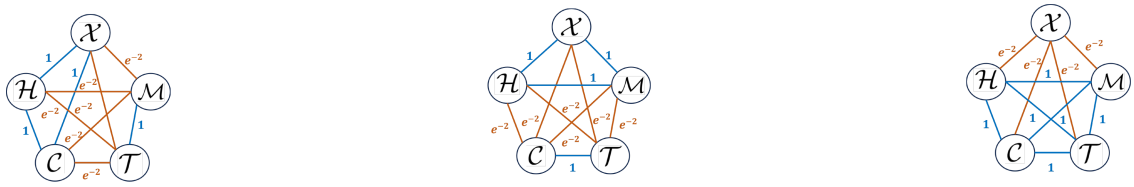


Fig. 5. Graphs in left, middle, and, right are for the three glioma tumor tissue samples shown in Fig. 1(a), (b), and (c) respectively.

values). This WSI sample has a reference ground-truth binary vector B_r (provided by the expert annotators in the dataset). So, one can create a complete reference graph G_r from B_r in the same manner as G is built from B.

The goal of the training process is to minimize the structural difference between the two graphs G and G_r. This difference is calculated using spectral decomposition of graphs [42]. Let A and A_r be the adjacency matrices of G and, G_r respectively. Further, let D and D_r be the diagonal degree matrices of G and, G_r respectively. Then, we can write:

$$L = I - D^{-1/2} A D^{1/2}$$
$$L_r = I - D_r^{-1/2} A_r D_r^{1/2} \tag{10}$$

where, L and L_r are respective symmetric normalized Laplacian of the graphs, and I is the identity matrix. We then compute the eigenvalues (λ) of the above Laplacians, and, define the spectral graph loss as:

$$\mathcal{L}_G = \sum_{i=1}^{k} |\lambda_{(i)} - \lambda_{r(i)}|^2 \tag{11}$$

where k is given by the following equation:

$$k = min\left[\left\{\arg\min_q \left(\frac{\sum_{i=1}^{q}\lambda_i}{\sum_{i=1}^{b}\lambda_i}\right) > \phi\right\}, \left\{\arg\min_t \left(\frac{\sum_{i=1}^{t}\lambda_{r(i)}}{\sum_{i=1}^{b}\lambda_{r(i)}}\right) > \phi\right\}\right] \tag{12}$$

Here, ϕ is a threshold. We keep the top k eigenvalues that contain $(\phi \times 100)\%$ of the energy Two graphs, G and G_r are more similar if $\mathcal{L}_G$ is closer to 0. So, the goal is to minimize $\mathcal{L}_G$.

4) Composite Loss: We have formulated three different loss functions to make the model learn three different types of information. The total loss is designed as a linear combination of the three loss functions and can be hence expressed as:

$$\mathcal{L}_\mathcal{T} = \alpha\mathcal{L}_C + \beta\mathcal{L}_\mathcal{P} + \gamma\mathcal{L}_\mathcal{G} \tag{13}$$

The three constants, α, β and γ appearing in the above equation, are determined experimentally.

C. Training Phase

After the WSI images are processed as described in Section III-A, we have two sets of 2D image patches, one belonging to $20\times$ magnification and another belonging to $5x$ magnification. We train two models having exactly the same architecture as shown in Fig. 3 for each of these two sets. In each model, deep features are first extracted by a CNN, which are fed into b FCNs.

Algorithm 1: Training Process.

Input: n_l number of local image patches each denoted as I_i^l
Output: b-dimensional binary prediction vector B

```
while loss is not minimized do
    i ← 0
    while i < n_l do
        Extract local features F_i^l from I_i^l using a CNN
        Compute b-dimensional probability vectors and
          p_i^l by inputting F_i^l into b number of FCNs
          respectively
        Compute b-dimensional binary prediction
          vectors B_i^l from and p_i^l respectively
        Compute ML-WCE Loss from Eq. 3 using B_i^l
        Compute Conditional Probability Loss from
          Eq. 5 using B_i^l
        Compute Spectral Graph Loss from Eq. 11
          using B_i^l
        Backpropagate losses
        i ← i + 1
    end
end
return B
```

Each FCN is trained to predict the status of a specific biomarker using the ML-WCE loss. Output probabilities from each FCN are first concatenated and then binarized to produce a binary vector. This binary vector is further used in the conditional probability based loss and spectral graph loss. The losses are then back-propagated to the CNN and FCNs. The spectral graph loss is calculated per sample as done in standard neural network architectures [43], [44] whereas the conditional probability loss is calculated per batch (see Appendix). The entire training workflow for $20\times$ magnification, as shown in Fig. 3, is then repeated for $5\times$ magnification. The algorithm for $20\times$ magnification is given below, which is the same for $5\times$ magnification. The steps in the training phase are summarized in Algorithm 1.

D. Testing Phase

During the testing phase, we extract patches at the local and global magnification levels from a WSI image in exactly similar manner as during the training phase. We need a single prediction for each of the b bio-markers from a WSI sample (of a patient data). So, we first average the prediction probabilities of each local patch to obtain the prediction vector for the test WSI [36]. Likewise, we average the probabilities of each global

Algorithm 2: Testing Process.

Input: n_l number of local image patches each denoted as I_i^l, n_g number of global image patches each denoted as I_i^g

Output: b-dimensional binary prediction vector B

$i \leftarrow 0$
$j \leftarrow 0$
while $i < n_l$ *and* $j < n_g$ **do**
  Extract local and global features F_i^l and F_i^g from I_i^l and I_i^g respectively using a CNN
  Compute b-dimensional probability vectors p_i^l and p_i^g by inputting F_i^l and F_i^g into b number of trained FCNs respectively
  Compute b-dimensional binary prediction vectors B_i^l and B_i^g from and p_i^l and p_i^g respectively
  $i \leftarrow i + 1$
  $j \leftarrow j + 1$
end
Average all B_i^l and B_i^g to get B^l and B^g respectively
Perform soft voting on B^l and B^g to get B
return B

patch to obtain the probability vector for the same test WSI. We finally apply a weighted soft voting approach to combine the two results to get a single binary output probability vector [45]. The weights are set automatically using a modulated rank averaging technique calculated from the training accuracies of local and global level models [46]. The testing workflow is shown in Fig. 4, where, we predict a value of 1 or 0 for each bio-marker from a test WSI image. Algorithm 2 lists the steps for the testing phase.

IV. Experiments

A. Data Preparation

In order to predict the presence or absence of genetic bio-markers from WSI, an image to label correspondence is required. The Cancer Imaging Archive (TCIA) has Hematoxylin and Eosin (H&E) stained WSI images of patients from the TCGA-GBM [21] and TCGA-LGG [20] datasets. The bio-marker information is available from the cBioPortal repository [22]. We mapped the WSI images from TCIA with the genetic bio-marker information for the same patient. Our dataset can be accessed through this link. The details of the data preparation are now presented below.

1) Genomic Data Selection and Processing: Clinical and genomic information about the patients of TCGA-GBM and TCGA-LGG dataset are available from the TCGA PanCancer Atlas dataset in the cBioPortal [22]. We selected samples with information available on specifically five bio-markers, the names and different values for which are shown in Table I.

2) WSI Data Selection and Processing: WSI data is available in. svs format. Each file contains multiple zoom levels, capturing the H&E stained tissue from various magnifications. Due to the very high-resolution of WSI images, for example, $\sim 20000 \times 20000 \times 3$ in our dataset, rendering and subsequent training of deep learning algorithms on this data becomes unfeasible even with latest Graphics Processing Units (GPUs). Therefore, we adopted a patch-based approach inspired by the work of [36]. We divided each WSI into smaller patches of resolution $224 \times 224 \times 3$. We automatically selected relevant patches using a histogram based patch selection technique as explained in Section III-A. The overall patch generation and selection process is shown in Fig. 2.

3) Data Aggregation: We first removed duplicate patient entries and retained only the latest samples from both datasets. Since complete biomarker data for all patients were not available in the cBioPortal, we excluded samples missing more than two biomarkers. For those missing one or two biomarkers, neurologists inferred the status based on tumor type and available (biomarker) data. We then aggregated WSI data from TCIA and biomarker data from cBioPortal using matching subject IDs, ensuring all criteria were met. This resulted in a one-to-one mapping of 690 subjects—each with a WSI image and a five-element binary biomarker vector. The cohort includes 388 female and 302 male patients. Cancer types are distributed as: Glioblastoma Multiforme (275), Oligodendroglioma (226), and Astrocytoma (189). Biomarker counts are summarized in Table II.

4) Dataset Division: We randomly divided the cohort of 690 patients into three sets. Specifically, 483 patients (70%) were allocated for training, 69 (10%) for validation, and the remaining 138 (20%) were set aside for testing purposes.

TABLE II
Table Showing the Counts of Each Type of Genetic Marker

ID	Value: 1	Value: 0
$\mathcal{H}$	360	330
$\mathcal{C}$	133	557
$\mathcal{X}$	164	526
$\mathcal{M}$	472	218
$\mathcal{T}$	371	319

The values and ID are similar to those in the Table. I

TABLE III
Layer Architecture of FCN

Layer	Input shape	Output shape
Linear	batch size x 1280	batch size x 640
Dropout (0.1)	batch size x 640	batch size x 640
Linear	batch size x 640	batch size x 256
Dropout (0.1)	batch size x 256	batch size x 256
Linear	batch size x 256	batch size x 64
Dropout (0.1)	batch size x 64	batch size x 64
Linear	batch size x 64	batch size x 1

Each dropout layer drops 10% of the neurons randomly.

B. Experimental Settings

1) Network Architecture: We use Kimianet [47], a CNN pre-trained on large histopathology datasets, which outperforms other popular models like DenseNet [44] and EfficientNet [43]. The FCN comprises fully connected layers with dropout (10%) to reduce overfitting, followed by a sigmoid activation. The output has shape batch size $\times$ 1, representing the binary biomarker status. The FCN architecture is detailed in Table III. Our codes are available here.

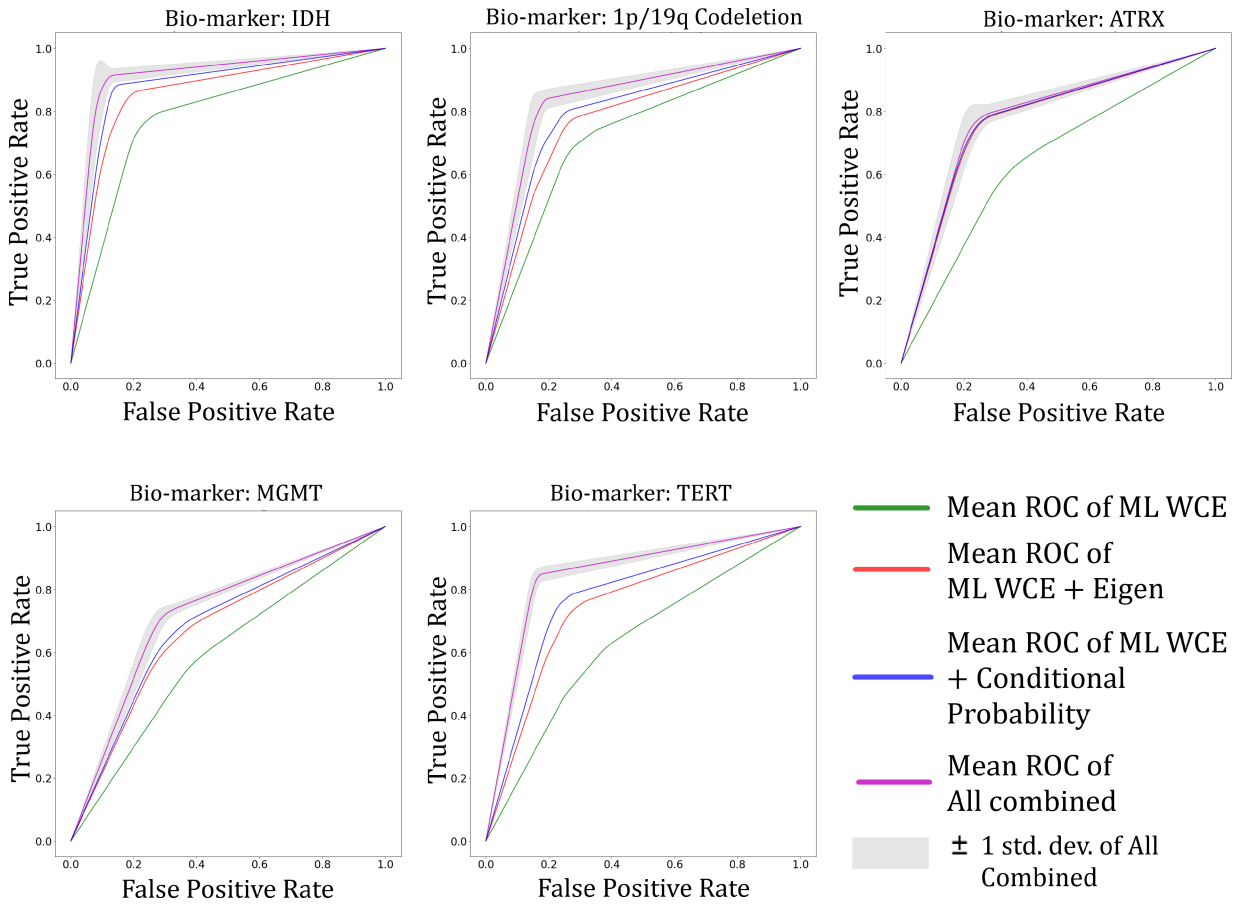


Fig. 6. The AUC for each bio-marker achieved by the models in shown in Table V. The gray bands represent 95% confidence interval for the AUC of our proposed model. The mean ROC of ML-WCE, ML-WCE + Spectral Loss, ML-WCE + Conditional Probability Loss and the mean ROC of all combined is shown with greenn, red, blue and magenta lines respectively.

TABLE IV
PERFORMANCE OF VARIOUS MAGNIFICATION PAIRS OF WSI

Magnification Pair	IDH	1p/19q	ATRX	MGMT	TERT
20× & 5×	0.913	0.845	0.783	0.720	0.849
20× & 10×	0.905	0.840	0.772	0.710	0.840
15× & 5×	0.900	0.835	0.775	0.715	0.842
20× & 15×	0.890	0.830	0.760	0.695	0.825
15× & 10×	0.889	0.825	0.758	0.705	0.828
10× & 5×	0.880	0.820	0.740	0.700	0.820

2) Impact of Magnification Levels: We experimented with different magnification levels and found out that using the highest (20x) and lowest (5x) magnification levels available in our dataset brought out the best results. The impact of various pairs of magnification of the WSI images are shown in Table IV.

3) Loss Function Parameters: We set $\phi = 0.9$ in ((12)), i.e., we keep the top k eigenvalues that have 90% of the energy. The other parameters ((13)) are used to control the weights (impact) of each component of the final loss function. We tried different combinations of α, β and γ values where each of them can have values belonging to the set $[0.1, 0.2, 0.3, \ldots, 1]$ similar to the procedure followed in [48]. After experimentation, we have found that $\alpha = 0.5, \beta = 0.2, \gamma = 0.3$ give optimum results.

4) Hyper Parameters: The learning rate and the batch size were respectively set to 3×10^{-4} and 256. We used the Adam optimizer for optimizing the weights and trained our model for 200 epochs. We further set $b = 5, m = 690, n_g = 200$ and $n_l = 2000$.

5) Computing Platform and Performance Metrics: All the computations are done using an HP Z640 Workstation with Intel Xeon 14-core Processor, 128 GB RAM and NVIDIA Titan RTX 24 GB Graphics processor. We employed PyTorch 1.9 [49] on an Ubuntu 20.04 platform. Considering the above hardware, the time taken to predict biomarkers from a single WSI sample is 10.0 minutes on average. This includes the time taken to split a WSI into its corresponding patches, running inference, and then combining the patch inference results to get the final prediction (for the whole sample). The average inference time for a single image patch is around 0.2 seconds.

Similar to most reported works on mutation prediction using histopathology images ([26], [50], [51]), we have employed AUC, Sensitivity, and Specificity as the performance measures in this work. We additionally computed 95% confidence intervals (CI) for the AUC.

V. RESULTS

During the training phase, individual image patches are utilized to train the model. For consistency, the same approach is

TABLE V
ABLATION STUDY: IMPACTS OF INDIVIDUAL (ML-WCE) LOSS, PAIRWISE (CONDITIONAL PROBABILITY BASED) LOSS, AND GROUPWISE (SPECTRAL GRAPH BASED) LOSS

Loss Function	IDH			1p/19q Codeletion			ATRX			MGMT			TERT			Mean
	AUC with 95% CI	Sens.	Spec.	AUC with 95% CI	Sens.	Spec.	AUC with 95% CI	Sens.	Spec.	AUC with 95% CI	Sens.	Spec.	AUC with 95% CI	Sens.	Spec.	AUC
ML-WCE	0.78 (0.729 - 0.812)	0.781	0.77	0.717 (0.636 - 0.771)	0.717	0.715	0.643 (0.556 - 0.715)	0.64	0.618	0.593 (0.5 - 0.654)	0.594	0.592	0.63 (0.518 - 0.763)	0.627	0.631	0.672
ML-WCE + Spectral Graph Loss	0.85 (0.800 - 0.912)	0.85	0.834	0.767 (0.706 - 0.859)	0.767	0.768	0.77 (0.726 - 0.824)	0.77	0.75	0.669 (0.587 - 0.751)	0.67	0.663	0.745 (0.677 - 0.822)	0.745	0.743	0.761
ML-WCE+ Conditional Probability Loss	0.88 (0.859 - 0.912)	0.87	0.88	0.793 (0.729 - 0.859)	0.792	0.79	0.772 (0.726 - 0.824)	0.773	0.764	0.682 (0.609 - 0.751)	0.68	0.65	0.773 (0.718 - 0.821)	0.778	0.77	0.78
Composite Loss = ML-WCE + Spectral + Conditional Probability Loss	**0.913 (0.879 - 0.932)**	**0.883**	**0.831**	**0.845 (0.812 - 0.878)**	**0.793**	**0.757**	**0.783 (0.735 - 0.815)**	**0.732**	**0.72**	**0.72 (0.696 - 0.752)**	**0.684**	**0.658**	**0.849 (0.829 - 0.867)**	**0.783**	**0.762**	**0.823**

followed during the testing phase. WSI of a subject is partitioned into smaller patches. The patch level probability vector outputs are aggregated and binarized into a WSI level binary vector, which serves as the final prediction.

A. Ablation Study

We have conducted an ablation study to reveal the impacts of individual loss functions on the prediction of each biomarker. We have calculated the AUC values along with 95% CI for each bio-marker. In the first model, we use only the ML-WCE loss function. This serves as a base model. In the second model, we combine ML-WCE loss, and, Spectral Graph Loss. The third model replaces Spectral Graph Loss with conditional probability loss. In the final model, all three loss functions are combined. As can be seen from Table V, the proposed model, which combines all the loss functions, achieves the highest AUC for all five bio-markers compared to the previous three combinations. We also show the mean AUC for each model. It is important to note that each loss function individually improves the mean AUC from the base ML-WCE model, which indicates that each loss function contributes significantly to the overall improvement. Fig. 6 further conveys this message in a qualitative manner. From Fig. 6 and Table V, we can see that MGMT prediction is challenging, as all the models performed below par in predicting MGMT. We created a confusion matrix as shown in Table VI to further corroborate that MGMT prediction is challenging due to high false negatives.

TABLE VI
CONFUSION MATRICES FOR FIVE BIOMARKERS

(a) IDH

	Pred +	Pred -
Actual +	64	8
Actual -	11	55

(b) 1p/19q

	Pred +	Pred -
Actual +	21	6
Actual -	27	84

(c) ATRX

	Pred +	Pred -
Actual +	24	9
Actual -	29	76

(d) MGMT

	Pred +	Pred -
Actual +	64	30
Actual -	15	29

(e) TERT

	Pred +	Pred -
Actual +	58	16
Actual -	15	49

'+' means positive and '-' means negative.

B. Comparison of Loss Functions

We have compared our loss function with other reported works on multi-label image prediction using different loss functions. Although, these works used nonmedical data, the modality is still 2D images like ours. Apart from a generic multi-label loss function, like Hamming loss [52], we re-implemented Asymmetric Loss [33], Distribution based Loss [32] and the best performing loss function - SpareMax Loss (SML) [53] from [34]. Table VII shows that our proposed approach outperforms the above works, thereby further reinforcing the need of developing a loss function, which can faithfully capture all the intricacies like individual, pairwise, groupwise behavior of the biomarkers.

C. State-of-The-Art Comparisons

In Table VIII, we have compared our work with state-of-the-art works on biomarker prediction. Note that most of these methods predicted only one biomarker. We first compare our multi-label prediction results, which is much more challenging against works that have done single biomarker prediction. Among them, most have attempted to predict the status of the IDH mutation because this is one of the most crucial biomarkers for the diagnosis of glioma tumors. It is true that Liu et al. [13] marginally outperformed us in IDH detection using WSI data. Nevertheless, it is important to mention that a binary prediction task is easier than that of a multi-label prediction. Kocak et al. [25] attempted to identify the 1p/19q codeletion status in which we outperform them in terms of AUC and specificity. We found two works that attempted MGMT status prediction. Out of these two methods, we outperform Saeed et al.'s work [23] by a huge margin. We lag slightly behind sensitivity of MGMT status prediction of [24], but outperform them on AUC, which represents the overall performance. Although Mora et al. [54] surpasses our ATRX detection, they incurred additional computational cost in model training for brain tumor segmentation followed by radiological features from it. Since, we have used WSI, no such extra computation is required. We did not find any

TABLE VII
COMPARISON OF VARIOUS LOSS FUNCTIONS FOR PREDICTION OF FIVE GENETIC MARKERS.

Method	IDH			1p/19q Codeletion			ATRX			MGMT			TERT			Mean
	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC
Hamming Loss [52]	0.764	0.713	0.723	0.717	0.683	0.672	0.643	0.617	0.623	0.598	0.572	0.581	0.738	0.724	0.731	0.692
SparseMax Loss (SML) [53] (2016)	0.847	0.84	0.791	0.798	0.748	0.723	0.73	0.71	0.684	0.631	0.632	0.62	0.729	0.687	0.664	0.747
Asymmetric Loss [33] (2019)	0.83	0.735	0.724	0.793	0.774	0.752	0.73	0.698	0.68	0.69	0.676	0.651	0.778	0.752	0.738	0.7642
Distribution Based Loss [32] (2020)	0.8	0.79	0.795	0.735	0.734	0.736	0.692	0.693	0.689	0.609	0.596	0.581	0.717	0.702	0.697	0.7106
Ours	**0.91**	**0.878**	**0.81**	**0.835**	**0.77**	**0.73**	**0.778**	**0.72**	**0.71**	**0.719**	**0.68**	**0.63**	**0.845**	**0.75**	**0.75**	**0.817**

TABLE VIII
COMPARISON WITH STATE-OF-THE-ART

Method	IDH			1p/19q Codeletion			ATRX			MGMT			TERT			Mean
Single biomarker prediction																
	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC
Liu et. al. [30] (2024)	0.886	0.793	0.755													-
Liu et. al. [13] (2020)	**0.927**	**0.889**	0.813	-												-
Liechty et. al. [29] (2022)	0.870	0.76	0.709													-
Kocak et. al. [25] (2019)	-			0.838	**0.877**	0.675	-									-
Faghani et. al. [24] (2023)	-									0.65	**0.712**	0.589	-			-
Saeed et. al [23] (2022)										0.53	-	-				-
Mora et. al. [54] (2023)	-						**0.831**	**0.772**	0.697	-						-
Ours	0.913	0.883	**0.831**	**0.845**	0.793	**0.757**	0.783	0.732	**0.72**	**0.72**	0.684	**0.658**	**0.849**	**0.783**	**0.762**	-
Two biomarkers prediction																
Zhou et. al. [26] (2019)	0.916	-	-	0.716	-	-	-	-	-	-	-	-	-	-	-	0.816
Nishikawa et al. [27] (2023)	0.814	-	-	0.771	-	-	-									0.7925
Tripathi and Bag [5] (2023)	**0.919**	**0.909**	**0.934**	**0.878**	**0.895**	**0.915**	-									**0.8985**
Ours	0.913	0.883	0.831	0.845	0.793	0.757	-									0.879
Five biomarkers prediction																
Le et. al. [12] (2025)	0.79	0.79	0.75	0.80	0.81	0.76	0.65	0.70	0.67	0.67	0.51	0.58	0.73	0.55	0.62	0.728
Xiao et. al. [55] (2023)	0.812	0.784	0.767	0.736	0.712	0.698	0.685	0.675	0.691	0.673	0.614	0.605	0.793	0.755	0.741	0.74
Weng et. al. [56] (2023)	0.913	0.854	0.817	0.823	0.807	0.813	0.772	0.746	0.712	0.697	0.641	0.623	0.824	0.795	0.786	0.8058
Densenet [44]	0.71	0.7	0.72	0.683	0.62	0.61	0.56	0.57	0.524	0.499	0.48	0.52	0.657	0.58	0.53	0.6218
Efficientnet (b0) [43]	0.739	0.74	0.71	0.631	0.68	0.623	0.6	0.64	0.59	0.557	0.58	0.547	0.588	0.64	0.6	0.623
Ours	**0.913**	**0.883**	**0.831**	**0.845**	**0.793**	**0.757**	**0.783**	**0.732**	**0.72**	**0.72**	**0.684**	**0.658**	**0.849**	**0.783**	**0.762**	**0.823**

AUC, Sensitivity (Sens.) and Specificity (Spec.) values are given, as per availability. Best values are highlighted in bold.

work that attempted to identify TERT mutation status, which is an important biomarker for tumor prognosis. It is natural to expect that the multi-label performance would not be able to exceed single biomarker prediction performance. But for proper diagnosis and prognosis of brain tumors, information on multiple biomarkers is collectively required [57], which renders the attempt to predict the status of just one biomarker unsuitable for a real clinical scenario.

We then list some works that have attempted to identify the status of two biomarkers - IDH and 1p/19q codeletion. We did not find any works that have tried to delve into simultaneous prediction of the status of other biomarkers like ATRX, MGMT and TERT. Zhou et al. [26] targeted prediction of two biomarkers - 1p/19q and IDH. Here, we lose in the IDH prediction extremely marginally; but, our overall mean is quite better in comparison. Nishikawa et al. [27] also attempted to predict the same set of biomarkers, we outperformed them by a big margin. Tripathi and Bag [5] obtained better results in both IDH and 1p/19q codeletion status prediction. However, their method is not suitable for prediction of other biomarkers, as discussed in Secction II. Also, they had to segment the tumor region first before attempting to identify biomarkers as only the tumor region has features related to the correct biomarker status of the patient while the rest of the healthy brain region has a different biomarker status and learning from the entire brain scan would thus, confuse the model. Segmenting the tumor required additional computational burden and processing times, which is not needed in WSI based prediction models like ours.

Finally, we compared our work on five-class multi label prediction with some state-of-the-art methods on multi-label prediction. We first compare our work with Le et. al. [12]. We show their results on the five biomarkers that we have worked upon. Although they have considered four additional biomarkers, our individual AUC for each biomarker is superior and shows that using the relationship information (pair-wise and group-wise) enhances the model performance. Also, we trained a single end-to-end network with five heads trained jointly, capturing individual, pairwise, and groupwise biomarker interactions, whereas, Le et. al. [12] used seperate networks to train each biomarker which is consumes more resource. We re-implemented Xiao et al.'s work [55] that performs a similar task of multi-label classification on our dataset and found that the performance is not up to the mark as they have focused on learning image features without paying attention to the label correlation. We also compared our work with another work that focus on label correlation [56] by re-implementing it on our dataset and found that it lags behind our results. We have further included comparisons with other baseline models, namely, DenseNet and EfficientNet, both of which are pre-trained with ImageNet [58] and fine-tuned only in the last FCN layer using a binary cross-entropy loss function. The results clearly establish the supremacy of our method. Overall, we can see that our

proposed method outperforms multi-label prediction models and also a few single label prediction models. We have also achieved the highest mean AUC value.

VI. Conclusion

In this paper, we simultaneously predicted five different genetic markers, i.e., IDH, 1p/19q codeletion status, ATRX, MGMT, and TERT, which are responsible for causing different types of brain tumors, from WSI data. Our solution harnesses different kinds of prevalent traits (individual, pairwise, and, groupwise behavior) in these biomarkers, which are explicitly reported in the scientific literature. We integrate conditional probability based loss capturing pairwise information, and, spectral graph based loss modeling group behavior with a multi-label weighted cross-entropy loss for individual traits (for the five genetic markers) in the form of a composite loss function. Experimental results demonstrate that state-of-the-art performance has been achieved on the benchmark. In the future, we plan to incorporate explainability into our proposed method as done in [59] so that medical practitioners can know the exact reasoning behind the predictions. This will make the model more trustworthy and transparent to the end users. We will further improve the prediction performance of MGMT and ATRX biomarkers by implementing multiscale attention, finer patch sampling for these specific biomarkers, and handling label noise [60]. Finally, one direction of future research will be to extend the present framework to predict additional biomarkers, such as, epidermal growth factor receptor (EGFR), tumor suppressor gene TP53, and CDKN2A for making the prognosis and treatment further comprehensive.

References

[1] Q. T. Ostrom et al., "The epidemiology of glioma in adults: A "state of the scienc," review," *Neuro Oncol.*, vol. 16, no. 7, pp. 896–913, 2014.

[2] W. L. Bi and R. Beroukhim, "Beating the odds: Extreme long-term survival with glioblastoma," *Neuro Oncol.*, vol. 16, no. 9, pp. 1159–1160, 2014.

[3] A. BS, A. B., H. RS, V. V., A. Mahadevan and H. J. Pandya, "Electromechanical characterization of human brain tissues: A potential biomarker for tumor delineation," *IEEE Trans. Biomed. Engg.*, vol. 69, no. 11, pp. 3484–3493, Nov. 2022.

[4] C. H. Suh, H. S. Kim, S. C. Jung, C. G. Choi, and S. J. Kim, "Imaging prediction of isocitrate dehydrogenase (IDH) mutation in patients with glioma: A systemic review and meta-analysis," *Eur. Radiol.*, vol. 29, pp. 745–758, 2019.

[5] P. C. Tripathi and S. Bag, "An attention-guided CNN framework for segmentation and grading of glioma using 3D MRI scans," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 20, no. 3, pp. 1890–1904, May/Jun. 2023.

[6] J. E. V. Vega and D. J. Brat, "Incorporating advances in molecular pathology into brain tumor diagnostics," *Adv. Anat. Pathol.*, vol. 25, no. 3, pp. 143–171, 2018.

[7] Z. Yi, L. Long, Y. Zeng, and Z. Liu, "Current advances and challenges in radiomics of brain tumors," *Front. Oncol.*, vol. 11, 2021, Art. no. 732196.

[8] J. Shao, J. Ma, Q. Zhang, W. Li, and C. Wang, "Predicting gene mutation status via artificial intelligence technologies based on multimodal integration (MMI) to advance precision oncology," in *Seminars in Cancer Biology*. Amsterdam, Netherlands: Elsevier, 2023, pp. 1–15.

[9] A. De, R. Mhatre, M. Tiwari, and A. S. Chowdhury, "Brain tumor classification from radiology and histopathology using deep features and graph convolutional network," in *Proc. Int. Conf. Pattern Recognit.*, 2022, pp. 4420–4426.

[10] L. Fu, Q. Peng, and L. Chai, "Predicting DNA methylation states with hybrid information based deep-learning model," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 17, no. 5, pp. 1721–1728, Sep./Oct. 2020.

[11] Y. Liu, A. Li, J. Liu, G. Meng, and M. Wang, "TSDLPP: A novel two-stage deep learning framework for prognosis prediction based on whole slide histopathological images," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 19, no. 4, pp. 2523–2532, Jul./Aug. 2022.

[12] M.-K. Le et al., "Glioma image-level and slide-level gene predictor (GLISP) for molecular diagnosis and predicting genetic events of adult diffuse glioma," *Bioengineering*, vol. 12, no. 1, 2024, Art. no. 12.

[13] S. Liu et al., "Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning," *Sci. Rep.*, vol. 10, no. 1, 2020, Art. no. 7733.

[14] R. B. Jenkins et al., "A t (1; 19)(q10; p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis of patients with oligodendroglioma," *Cancer Res.*, vol. 66, no. 20, pp. 9852–9861, 2006.

[15] M. Szylberg et al., "MGMT promoter methylation as a prognostic factor in primary glioblastoma: A single-institution observational study," *Biomedicines*, vol. 10, no. 8, 2022, Art. no. 2030.

[16] Y. Li et al., "Genotype prediction of ATRX mutation in lower-grade gliomas using an MRI radiomics signature," *Eur. J. Radiol.*, vol. 28, pp. 2960–2968, 2018.

[17] N. Olympios, V. Gilard, F. Marguet, F. Clatot, F. D. Fiore, and M. Fontanilles, "TERT promoter alterations in glioblastoma: A systematic review," *Cancers*, vol. 13, no. 5, 2021, Art. no. 1147.

[18] P. Śledzińska, M. G. Bebyn, J. Furtak, J. Kowalewski, and M. A. Lewandowska, "Prognostic and predictive biomarkers in gliomas," *Int. J. Mol. Sci.*, vol. 22, no. 19, 2021, Art. no. 10373.

[19] A. Arcella et al., "Dissecting molecular features of gliomas: Genetic loci and validated biomarkers," *Int. J. Mol. Sci.*, vol. 21, no. 2, 2020, Art. no. 685.

[20] N. Pedano et al., "The cancer genome atlas low grade glioma collection (TCGA-LGG)," Canc. Imag. Arch., 2016.

[21] L. Scarpace et al., "The cancer genome atlas glioblastoma multiforme collection (TCGA-GbM)," *Cancer Imag. Arch.*, 2016.

[22] E. Cerami et al., "The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data," *Cancer Discov.*, vol. 2, no. 5, pp. 401–404, 2012.

[23] N. Saeed, S. Hardan, K. Abutalip, and M. Yaqub, "Is it possible to predict MGMT promoter methylation from brain tumor MRI scans using deep learning models?," in *Proc. Int. Conf. Med. Imag. Deep Learn.*, 2022, pp. 1005–1018.

[24] S. Faghani et al., "A comparison of three different deep learning-based models to predict the MGMT promoter methylation status in glioblastoma using brain MRI," *J. Digit. Imag.*, vol. 36, no. 3, pp. 837–846, 2023.

[25] B. Kocak et al., "Radiogenomics of lower-grade gliomas: Machine learning–based MRI texture analysis for predicting 1p/19q codeletion status," *Eur. Radiol.*, vol. 30, pp. 877–886, 2020.

[26] H. Zhou et al., "Machine learning reveals multimodal MRI patterns predictive of isocitrate dehydrogenase and 1p/19q status in diffuse low-and high-grade gliomas," *Neuro Oncol.*, vol. 142, pp. 299–307, 2019.

[27] T. Nishikawa et al., "Easy-to-use machine learning system for the prediction of IDH mutation and 1p/19q codeletion using MRI images of adult-type diffuse gliomas," *Brain Tumor Pathol.*, vol. 40, no. 2, pp. 85–92, 2023.

[28] K. Masui et al., "Elevated TERT expression in TERT-wildtype adult diffuse gliomas: Histological evaluation with a novel TERT-specific antibody," *BioMed. Res. Int.*, vol. 2018, no. 1, 2018, Art. no. 7945845.

[29] B. Liechty et al., "Machine learning can aid in prediction of IDH mutation from h&e-stained histology slides in infiltrating gliomas," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 22623.

[30] X. Liu et al., "Multi-scale feature fusion for prediction of IDH1 mutations in glioma histopathological images," *Comput. Methods Programs Biomed.*, vol. 248, 2024, Art. no. 108116.

[31] D. Wang et al., "Automated machine-learning framework integrating histopathological and radiological information for predicting IDH1 mutation status in glioma," *Front. Bioinform.*, vol. 1, 2021, Art. no. 718697.

[32] T. Wu, Q. Huang, Z. Liu, Y. Wang, and D. Lin, "Distribution-balanced loss for multi-label classification in long-tailed datasets," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 162–178.

[33] T. Ridnik et al., "Asymmetric loss for multi-label classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 82–91.

[34] H. Yessou, G. Sumbul, and B. Demir, "A comparative study of deep learning loss functions for multi-label remote sensing image classification," in *Proc. IEEE Int. Geosc. Remote Sens. Symp.*, 2020, pp. 1349–1352.

[35] X. Xie et al., "A survey on incorporating domain knowledge into deep learning for medical image analysis," *Med. Image Anal.*, vol. 69, 2021, Art. no. 101985.

[36] N. Dimitriou, O. Arandjelović, and P. D. Caie, "Deep learning for whole slide image analysis: An overview," *Front. Med.*, vol. 6, 2019, Art. no. 264.
[37] A. Patil et al., "Efficient quality control of whole slide pathology images with human-in-the-loop training," *J. Pathol. Inform.*, vol. 14, 2023, Art. no. 100306.
[38] B. Li, Y. Li, and K. W. Eliceiri, "Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 14318–14328.
[39] A. Ashtaiwi, "Optimal histopathological magnification factors for deep learning-based breast cancer prediction," *Appl. Syst. Innov.*, vol. 5, no. 5, 2022, Art. no. 87.
[40] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, pp. 19–67, 2005.
[41] H. Arita et al., "Upregulating mutations in the TERT promoter commonly occur in adult malignant gliomas and are strongly associated with total 1p19q loss," *Acta Neuropathologica*, vol. 126, pp. 267–276, 2013.
[42] B. Nica, *A Brief Introduction to Spectral Graph Theory, Ser. EMS Textbooks in Mathematics*. Helsinki, Finland: European Mathematical Society, 2018.
[43] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.
[44] G. Huang et al., "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.
[45] D. Kumar and U. Batra, "Breast cancer histopathology image classification using soft voting classifier," in *Proc. 3rd Int. Conf. Comput. Inf. Netw.*, 2021, pp. 619–631.
[46] A. De and A. S. Chowdhury, "DTI based alzheimer's disease classification with rank modulated fusion of CNNs and random forest," *Expert. Syst. Appl.*, vol. 169, 2021, Art. no. 114338.
[47] A. Riasatian et al., "Fine-tuning and training of densenet for histopathology image representation using TCGA diagnostic slides," *Med. Image Anal.*, vol. 70, 2021, Art. no. 102032.
[48] S. Rajaraman, G. Zamzmi, and S. K. Antani, "Novel loss functions for ensemble-based medical image classification," *Plos One*, vol. 16, no. 12, 2021, Art. no. e0261307.
[49] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, vol. 32, pp. 8024–8035.
[50] X. Wang et al., "Prediction of BRCA gene mutation in breast cancer based on deep learning and histopathology images," *Front. Genet.*, vol. 12, 2021, Art. no. 661109.
[51] N. Tomita et al., "Predicting oncogene mutations of lung cancer using deep learning and histopathologic features on whole-slide images," *Transl. Oncol.*, vol. 24, 2022, Art. no. 101494.
[52] E. Frank and M. Hall, "A simple approach to ordinal classification," in *Proc. Eur. Conf. Mach. Learn.*, 2001, pp. 145–156.
[53] A. Martins and R. Astudillo, "From softmax to sparsemax: A sparse model of attention and multi-label classification," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1614–1623.
[54] N. Mora et al., "Comparison of MRI sequences to predict ATRX status using radiomics-based machine learning," *Diagnostics*, vol. 13, no. 13, 2023, Art. no. 2216.
[55] J. Xiao, Y. Bai, A. Yuille, and Z. Zhou, "Delving into masked autoencoders for multi-label thorax disease classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2023, pp. 3588–3600.
[56] W. Weng, B. Wei, W. Ke, Y. Fan, J. Wang, and Y. Li, "Learning label-specific features with global and local label correlation for multi-label classification," *Appl. Intell.*, vol. 53, no. 3, pp. 3017–3033, 2023.
[57] C. Horbinski et al., "The medical necessity of advanced molecular testing in the diagnosis and treatment of brain tumor patients," *Neuro Oncol.*, vol. 21, no. 12, pp. 1498–1508, 2019.
[58] O. Russakovsky et al., "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
[59] F. H. Yagin et al., "Explainable artificial intelligence model for identifying COVID-19 gene biomarkers," *Comput. Methods Programs Biomed.*, vol. 154, 2023, Art. no. 106619.
[60] A. Jurgas, M. Wodzinski, M. D'Amato, J. van der Laak, M. Atzori, and H. Müller, "Improving quality control of whole slide images by explicit artifact augmentation," *Sci. Rep.*, vol. 14, no. 1, 2024, Art. no. 17847.

Arijit De (Student Member, IEEE) received the MTech in computer science and engineering from Jadavpur University, Kolkata, India. He is currently working toward the PhD degree with Jadavpur University. His research interests include biomedical imaging, clinical neuroimaging, and solving medical imaging problems with computer vision.

Aritro Santra received the BTech degree in electronics and communication engineering from the Manipal Institute of Technology, Manipal, India, in 2023. He is currently working as a Salesforce Application Developer with Deloitte USI, Kolkata, India. His research interests include deep learning for biomedical imaging, bioinformatics with deep learning.

Mona Tiwari received the MBBS degree from MGM Medical College, Indore, India, the DMRD degree in medical radiodiagnosis from SMS Medical College, Jaipur, India, and the FRCR degree from the Royal College of Radiology, London, U.K. She is currently the radiological co-investigator (senior consultant radiologist) with the team working with the Institute of Neurosciences, Kolkata. She is the head of the radiology department and focuses on the diagnosis of brain tumors and other neurological disorders.

Ananda S. Chowdhury (Senior Member, IEEE) received the PhD degree in computer science from the University of Georgia, Athens, GA, USA, in 2007. He was a postdoctoral fellow with the Department of Radiology and Imaging Sciences, National Institutes of Health, Bethesda, MD, USA, from 2007 to 2008. He is currently a professor with the Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India, where he leads the Imaging, Vision, and Pattern Recognition Group. He has authored or co-authored more than 90 papers in leading international journals and conferences, in addition to a monograph in the Advances in Computer Vision and Pattern Recognition series of Springer. His research interests include computer vision, pattern recognition, biomedical image/signal processing, and multimedia analysis. Dr. Chowdhury is a member of the IAPR-TC on Graph-Based Representations. He serves as an associate editor for *IEEE Transactions on Image Processing* and an area editor for *Pattern Recognition Letters*. His Erdős number is two.