Comprehensive Genomic Subtyping of Glioma Using Semi-Supervised Multi-Task Deep Learning on Multimodal MRI

PRIYANKA TUPE-WAGHMARE\textsuperscript{1,2}, (Senior Member, IEEE), PIYUSH MALPURE\textsuperscript{1,2}, KETAN KOTECHA\textsuperscript{1}, MANISH BENIWAL\textsuperscript{3}, VANI SANTOSH\textsuperscript{3}, JITENDER SAINI\textsuperscript{3}, AND MADHURA INGALHALIKAR\textsuperscript{2}

\textsuperscript{1}Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, Maharashtra 412115, India
\textsuperscript{2}Symbiosis Centre for Medical Image Analysis, Symbiosis International (Deemed University), Pune, Maharashtra 412115, India
\textsuperscript{3}National Institute of Mental Health and Neurosciences, Bengaluru 560029, India

Corresponding author: Madhura Ingalhalikar (head@scmia.edu.in)

This work was supported in part by the Research Support Fund of Symbiosis International (Deemed University), Pune, India, and in part by the Core Research Grant through the Science and Engineering Research Board (SERB) Department of Science and Technology (DST) under Grant CRG/2020/003258.

This work involved human subjects in its research. Approval of all the ethical and experimental procedures was granted by the Independent Ethics Committee (Reference: SIU/IEC/231).

\textbf{ABSTRACT} High grade glioma (HGG) are the most common, highly infiltrative brain tumors usually with a grim outcome with low survival. Recent comprehensive genomic profiling has greatly elucidated the molecular markers in gliomas that include the mutations in isocitrate dehydrogenase (IDH), 1p/19q co-deletion and alterations in O6 methyl-guanine methyltransferase (MGMT). Prognosis of these markers can support early clinical decision making that can consequently improve the survival outcomes. Multimodal MRI based phenotypic tools such as radiomics based multi-variate models and state-of-art convolutional neural networks (CNNs) have shown promise in identifying these genotypes. However, current techniques do not facilitate comprehensive genomic profiling of the HGG as these are focused only on a single mutation. Moreover, the models are trained on small datasets and cannot employ unlabeled data, which is abundant as pathological labelling is invasive, expensive and inaccessible in many places. In this work, we build a semi-supervised hierarchical multi-task model that can incorporate unlabeled glioma data and learn to predict multiple molecular markers simultaneously. Our framework employs the latent space from an encoder to incorporate the unlabeled data while the hierarchical multi-task model accounts for the similarity between tasks and utilizes the shared information, resulting in inductive learning that facilitates precise delineation of IDH, MGMT, 1p/19q and grade. We applied our framework to 120 labeled, 149 semi-labeled and 48 unlabeled data using T1-contrast enhanced, T2 and FLAIR images and illustrate that our model performs with an average test accuracy of 82.35\% and verified the results using task wise and modality wise ablation analysis. Moreover, the class activation maps computed from each local task branch provide clinical interpretability.

\textbf{INDEX TERMS} Autoencoder, deep learning, glioma, machine learning, MRI, molecular markers, radiogenomics.

\textbf{I. INTRODUCTION}
High grade glioma (HGG) are prevalent neoplasms emerging from the glial cells of the central nervous system (CNS). These tumors are associated with poor prognosis and very low survival rates (median 15 months) \cite{1,2,3}.

Early prognosis and timely management of these tumors is essential as the pathogenesis is highly complicated and outcomes are often poor despite intensive multimodal management involving surgical resection, adjuvant radiotherapy and chemotherapy \cite{4}.

In the past few years, with the emergence of molecular profiling, there has been a paradigm shift in the diagnosis of glioma, indicating that both histology and molecular
classification are crucial. This has been highlighted in the deliberations from 2016 World Health Organization [5] and updated by the Consortium to Inform Molecular and Practical Approaches to CNS Tumor Taxonomy (cIMPACT-NOW) where each categorization is now based upon an integrated histology and molecular profile classification [6]. Molecular factors such as presence or absence of isocitrate dehydrogenase (IDH) 1 and 2 genotypes are crucial as these have been directly linked to patient survival, with a consequent clinical and prognostic impact [7], [7]. IDH mutant glioma are further categorized as astrocytic and oligodendroglial types based on 1p/19q codeletion status. Here 1p is related to the loss of the short arm of chromosome 1 and 19q refers to the loss of the long arm of chromosome 19 and the co-deletion status is a prognostic indicator for therapeutic response to chemotherapy and chemoradiotherapy [9]–[11].

Another important basis for categorization is the methylation of the O6 methyl-guanine methyl transferase (MGMT) gene promoter. MGMT methylation results in increased efficacy of alkylating agents and therefore is a predictive factor for chemotherapy response and consequently the overall survival. Current standard care to delineate tumor type includes histopathology and immunohistochemistry [12], [13] which requires a tissue specimen that is extracted using an invasive procedure such as complete tumor resection or biopsy. However, resecting tumors from deep and eloquent locations is complicated and comes with a risk of morbidity. Moreover, longitudinal assessment over the course of treatment using resected tissue sample is infeasible [14]. To this end, MRI based phenotypic markers have potential to assist in non-invasive pre-operative prognosis of the mutational status that can support a customized treatment plan catering to each patient in the initial stage of the disease.

Radio-genomics includes quantitative analysis of radiophenotypic features extracted from radiological images to identify the genotype of the tumor. Intensity, shape, and textural features, together termed as radiomics, extracted from T1-CE enhancement and FLAIR hyperintense regions in addition to tumor volume and location have been commonly employed to identify the IDH mutational status [15], [16], 1p/19q codeletion [17] as well as MGMT methylation status [18]. With recent advancements in artificial intelligence (AI), automated learning of discriminative patterns in multi-modal MRI has become state-of-art. Models based on Convolutional Neural Networks (CNNs) have illustrated remarkable performance in multimodal MRI based image classification including grade and IDH genotype prediction [19] and MGMT classification [20]. CNNs have the capability to learn the image-based features automatically, alleviating the time-consuming pre-processing and feature extraction steps required in radiomics [21]. These auto-engineered features from the CNNs with a complete end-to-end optimization is an attractive and efficient model for predictive tasks. Nevertheless, current CNN based models only learn the underlying patterns to delineate a single subtype (for example, IDH mutant or wildtype). However, with the new updates from cIMPACT-NOW it is becoming more evident that diagnosis based on multiple molecular characteristics is crucial for diagnostic categorization. For example, low grade, IDH mutation and 1p/19q co-deletion are required to define oligodendroglioma while the cIMPACT-NOW update now states that grade IV and IDH wildtype status both are necessary to define glioblastoma. Therefore, multiple categorizations need to be performed simultaneously to facilitate glioma subtyping. Besides, techniques mentioned in the previous studies have ignored subjects having partial or fully missing labels (ground truth molecular categorization using immunohistochemistry). Generally, the proportion of subjects who do not undergo immunohistochemistry is higher as the procedure is not yet accessible in all parts of the world. Including these subjects is crucial for learning the variability in the data that can train the model in a robust fashion.

In this paper, we address both the issues (1) simultaneous multiple classifications and (2) the use of unlabeled data by harnessing the power of a semi-supervised multi-task artificial intelligence-based framework. We learn the underlying common as well as discrete patterns using CNN while facilitating tumor categorization using grade, IDH, 1p/19q and MGMT on labeled as well as unlabeled data. Finally, for interpretability of the classifier, for each task, saliency maps are computed using a novel high resolution class activation maps-based technique (HR-CAM). The results of this multi-task model are compared to multilabel and multiclass models.

II. METHODOLOGY
A. SUBJECT RECRUITMENT

The dataset consisted of 375 high grade glioma cases with 3 modalities – T1-contrast enhanced T1-CE, T2-weighted and FLAIR. 150 subjects were obtained from The Cancer Imaging Archive (TCIA) while 225 were scanned at National Institute of Mental Health and Neurosciences, Bengaluru, India (NIMHANS) (shown in TABLE 1). The retrospective study was approved by the institutional ethics committee and the requirement to obtain informed consent was waived. The patients had also undergone surgical resection, standard post-surgical care. After the quality check of the MRI, availability of multimodal data and discussion with the clinical experts

| Source                        | Data count | Final Count used |
|-------------------------------|------------|------------------|
| The Cancer Genome Atlas (TCGA) | 150        | 150              |
| Local Dataset from National Institute of Mental Health and Neurosciences (NIMHANS) | 225        | 157              |
| Total                         | 375        | 307              |
(JS- has 15 years of experience in clinical neuroimaging and VS- has 25 years of experience in clinical neuropathology) only 307 out of 375 cases were included in the study. 48 subjects out of the 307 were completely unlabeled, 269 with IDH status available, 94 with MGMT status available, 146 with 1p/19q co-deletion status available and 269 with grade information. TABLE 2 provides the details.

**B. DATA ACQUISITION**

The local data were scanned on either Philips 3T Ingenia or Siemens 1.5T Aira scanners. The study involved multiple sequences. However since our focus is on T1-CE, FLAIR and T2 contrast, the acquisition parameters of these are given: (1) 3D T1-CE: 1.5T-TR/TE = 1800-2200/2.6-2.9 ms; 3T-TR/TE = 1900-2200/2.3-2.4 ms in 0.5mmx0.5 mmx5 (2) FLAIR: 1.5T - TR/TE = 7500 - 11000/8.2 - 9.7 ms; 3T-TR/TE = 9000/8.1-9.4 ms in plane resolution 0.5mmx0.5 mmx5.(3) T2: 1.5T- TR/TE = 4900-6700/89-99 ms and 3T- TR/TE = 5200-6700/80-99 ms in 0.5mmx0.5 mmx5. The inversion times (TI) = 1800 ms. The images acquired from TCIA were scanned of multiple scanners with multiple protocols. The details can be found in Bakas et al. [22].

**C. PRE-PROCESSING**

The preliminary step of processing is the conversion of images acquired in DICOM® to NIFTI format followed by manual QA and elimination of subjects with distortions and motion. Brain extraction was performed on the volumes of all the modalities of each patient using FSL’s brain extraction tool (BET) (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET/UserGuide). All the modalities (T2W and FLAIR) were registered to 3D T1CE with rigid transformation using the Advanced Normalization Tools (ANTs) tool [23]. This was followed by automated segmentation of the tumoral regions into contrast enhanced tumor, edema and necrosis using a U-net based Deepmedic model that was trained and validated on 256 subjects from BRATS dataset [24]–[26]. The segmentations were then manually corrected using ITK-SNAP (http://www.itksnap.org) by an expert annotator and checked by a radiologist.

The labeled dataset (fully labelled as well as semi-labelled) was split into two cohorts: training cohort consisting of 80% (96 subjects) and testing cohort consisting of 20% (24 subjects) of the data. The semi-labeled data, consisting of 149 subjects, was combined with the training cohort making the data count 245. The unlabeled dataset, of 48 subjects, was used as a training cohort to train the convolutional auto-encoder (CA). The model was applied on a boxed region around the tumor for multiple 2D-axial slices. Identical slices from each modality were stacked together and a 3-channel image was employed. For each subject, 15 slices were taken from the training cohort and 6 slices from the testing cohort.

**D. MODEL DESIGN**

We built a semi-supervised hierarchical multi-task learning framework specifically to delineate numerous glioma genotypes and grade based on multi-modal MRI (T1- contrast enhance (T1-CE), T2-weighted and FLAIR images).

Multi-task learning is an approach to learn multiple tasks while accounting for similarity of tasks enabling utilization of shared information. This results in an inductive learning framework that ultimately can improve the accuracy of the model. Here we propose to employ a novel semi-supervised multi-task architecture to leverage the huge unlabeled and semi-labeled dataset available in the domain. First, we train through a convolutional auto-encoder (CA) through
multimodal MRI images. The CA consisted of an encoder having 16 layers, containing four blocks with two convolution layers followed by max-pooling and batch normalization while the decoder consisted of four blocks with two transposed convolution layers followed by batch normalization. The concatenation layer merges the features from the CNN and CA. This is then followed by a global average pooling block represented by green and two dense layers. The last neuron is a sigmoid. The latent space features from the auto-encoder are then concatenated with the CNN architectures’ layer, late in the model (see Fig. 1). CAs do not require labeled data and can capture important features with structural similarities and therefore are critical in our framework.

1p/19q codeletion occurs incase of IDH mutation hence IDH wildtype and 1p/19q codeletion are kept mutually exclusive.

For the CNN, a shared representation is concatenated with the latent representation from the trained encoder that is then passed to task-specific outheads to compute the outputs from each task. The loss functions for our model are given as below:

\[
H(y) = -\frac{1}{N} \sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)
\]  

(1)

where, \( y \) is the true value of the label, \( \hat{y} \) = predicted value, \( N \) = Total number of subjects and \( H(y) \) is the cross-entropy loss.

\[
C(y, m) = H(y) \times (if(y_i = m))
\]  

(2)

where, \( C \) is masked cross-entropy, \( y \) = True value, \( m \) = mask value. The final loss function for the multi-task model is computed using equation (3).

\[
L(y, m) = C_{\text{Grade}}(y^{\text{Grade}}, m) + C_{IDH}(y^{IDH}, m) + C_{\text{1p19q}}(y^{\text{1p19q}}, m) + C_{\text{MGMT}}(y^{\text{MGMT}}, m)
\]  

(3)

where \( C \) is masked cross-entropy, with the mask value \( m \) set to \(-1\) when the label is not available. The true labels for missing labels are also set to \(-1\) such that the loss will be 0 for missing labels and the weights won’t get updated. The weight given to all the tasks is equal, however can always be changed based on the importance of the task.

As seen in Fig. 1 the 1p/19q co-deletion is a hierarchical task with the parent being IDH mutation prediction, as has been illustrated in WHO 2016 glioma characterization. Basically, the co-deletion can occur only in cases of mutated IDH. We introduce the hierarchy in our framework based on the work by Salakhutdinov et al. [27] where the correlation between high-level features and low-level generic features are shared between the hierarchical classes. The IDH mutation prediction neuron connected to the pre-prediction layer of the 1p/19q acts as a weighted bias to correlate the features during the training that facilitates the hierarchy.

E. COMPARITIVE MODELS

We compared multitask models with other models that can incorporate multiple labels at the same time. To this end, multi-class and multi-label models were used. Multi-class model with a base CNN of ResNet50 [28] was trained with 12 classes(G3_MGMT_IDH_MT_1pCo,G3_MGMT_IDH_MT_1pNonCo,G3_MGMT_IDH_WT_1pCo,G3_MGMT_IDH_WT_1pNonCo,G4_MGMT_IDH_MT_1pCo,G4_MGMT_IDH_MT_1pNonCo,G4_MGMT_IDH_WT_1pCo,G4_MGMT_IDH_WT_1pNonCo,G4_NMGMT_IDH_MT_1pCo,G4_NMGMT_IDH_MT_1pNonCo) similar to multitask classification label.

F. TRAINING AND TESTING

Training of all the 3 models was performed on Nvidia GPU GTX 1060 with 6 GB of memory. CA was trained for 500 epochs with the learning rate kept constant at 1e-4 using a dice loss.

For the base CNN, ResNet50 with pre-trained weights on the image-net dataset was employed and further trained on the tumor dataset. The multitask outheads considered the concatenated input from ResNet50 and latent space from the encoder. The complete multi-step training was performed for 425 epochs. The learning rate was gradually varied from 0.003 to 3 × 10e-7. The multi-label and multi-class models were also trained using multi-step training for 300 epochs with learning rate varying gradually from 0.003 to 3 × 10e-7. To prevent potential overfitting and to provide more samples to the model, image augmentation was performed to increase the data size by flipping the images vertically and horizontally, rotating them at an angle, random vertical and horizontal shifts, random zoom and random sheering. Gaining deeper insight into the interrelation between the tasks is crucial for clinical interpretability. To this end we have performed the task wise ablation analysis on Grade, IDH mutation status, 1p/19q codeletion status and MGMT promoter methylation status. Results for – all the tasks, skipping one task at a time and skipping two and three tasks were noted. We also have performed the modality wise ablation analysis on T1-CE, FLAIR and T2W images. Furthermore, we have calculated the accuracy of the model for the individual modalities. High Resolution Class activation maps (HR-CAMs) were computed from the final layer of each outhead task to attain model interpretability.

G. NETWORK INTERPRETABILITY

In this study, we have employed the high-resolution class activation maps (HR-CAMs) by Shinde et al. [29] to demonstrate the interpretability of the model. This technique facilitates enhanced visual explainability to the CNN models by extracting feature information from each layer that can localize abnormalities with greater details in original image resolution. HR-CAMs has demonstrated enhanced performance compared to CAMs [30] and Gradient CAMs [31] where the feature maps were extracted only from a single layer of the CNN and has been rigorously validated on several datasets.
In our study, four convolution layers from the branch of each task were considered to obtain the HR-CAMs as shown in Fig. 1. It can be seen that HR-CAMs of each task is focusing on different areas of the tumor where red color indicates the most focused area by the model and blue being the least focused area.

III. RESULTS

Our framework was first compared to multi-label and multi-class classifiers. The average training and test accuracies in our case were higher as shown in TABLE 3. Further, for our multi-task classifier, we computed the test accuracy, sensitivity, specificity, precision and the F1 score and is given in TABLE 4. We trained the CNN based semi-supervised multitask model using unlabeled data with 93.04 % accuracy for 1p/19q codeletion prediction, the accuracy of 82.51 % and 84.34 % for grade and MGMT status prediction. Training accuracy for predicting IDH mutation status was 83.12 %. For the labelled data, we achieved training accuracy of 94.58 %, 83.57 %, 70.33 % and 83.57 % for prediction of 1p/19q, grade, MGMT and IDH mutation, respectively. Test accuracy of 88.73 % (1p/19q), 79.68 % (grade), 80.96 % (MGMT) and 80.04 % (IDH) is achieved for unlabeled data whereas for the labelled data test accuracy of 89.55 % (1p/19q), 73.81 % (grade), 61.15 % (MGMT) and 74.85 % (IDH) is achieved. The accuracies were compared with the multitask classifier without using the unlabeled data by removing the latent space concatenation. It can be observed that employing unlabeled data boosted the classification accuracy significantly. The results for task wise ablation analysis by removing single task, two tasks and three tasks are shown in TABLE 5. A superior average accuracy is demonstrated for the analysis considering grade, IDH mutation and 1p/19q codeletion. Modality wise ablation analysis is depicted in TABLE 6 where highest accuracy is achieved for 1p/19q considering T1-CE and FLAIR images. The train-test results for
TABLE 3. Comparative results between multi-class, Multi-label and multi-task model.

| Model                                      | Average Training Accuracy | Average Test Accuracy | Average Precision | Average F1-score |
|-------------------------------------------|---------------------------|-----------------------|-------------------|------------------|
| Multi-Class                               | 64.08                     | 50.04                 | 0.59              | 0.48             |
| Multi-Label                               | 82.43                     | 60.25                 | 0.64              | 0.58             |
| Multi-task without autoencoder            | 81.94                     | 74.84                 | 0.69              | 0.63             |
| Multi-task with autoencoder              | 85.75                     | 82.35                 | 0.82              | 0.8              |

**FIGURE 2.** ROC curves for each task for the proposed model.

T1-CE, FLAIR and T2 are mentioned in TABLE 7 where we achieved 92.12% train accuracy and 86.92% test accuracy for 1p/19q codeletion with FLAIR modality. The receiver operating characteristics (ROC) for our method are displayed in Fig. 2. The heatmaps for each outhead task are displayed in Fig. 3. The first row depicts grade 3, IDH wildtype, 1p/19q non-codeleted without MGMT methylation while row 2 is an example of grade 4, IDH wildtype, without 1p/19q co-deletion and MGMT methylation. Third row depicts a grade 4 tumor which is IDH wildtype, MGMT methylated and without 1p/19q codeletion and fourth row represents a heatmap of grade 3 IDH mutated, 1p/19q codeleted and MGMT methylated tumor. The heatmaps are computed from the final layer and thus provide a gross representation of the areas attributing to the classification.

**IV. DISCUSSION**

This work developed a single unified technique using a semi-supervised hierarchical multi-task model that could predict multiple molecular biomarkers in high grade glioma using multi-modal pre-operative MRI. The model not only accounted for the unlabeled data but also could facilitate simultaneous predictions for grade, IDH, 1p/19q and MGMT. Comparisons with analogous multi-class and multi-label models illustrated that our proposed network performed with superior accuracy while the heatmaps for each task could illustrate that the model consistently focused on the tumoral region.

Multi-task models have been employed in numerous medical imaging applications. Choi et al. [32] illustrated an accuracy of 86.83% using a standard ResNet model,
FIGURE 3. Class Activation Maps computed from four subjects illustrate that the model focuses on the tumoral area for classification. First column shows the input image and heatmaps obtained by HR-CAM method for grade, IDH mutation, MGMT methylation and 1p/19q codeletion are shown by column 2 to column 4 respectively. First row represents – G3, IDH WT, 1p/19q non-codeleted, MGMT methylated tumor. Second row represents- G4, IDH WT, 1p/19q non-codeleted, MGMT unmethylated tumor. Third row represents – G4, IDH WT, 1p/19q non codeleted, MGMT methylated tumor and the last row represents G3, IDH MT, 1p/19q codeleted and MGMT methylated tumor.

Yogananda C. et al. in 2019 demonstrated 97.14% accuracy whereas Pasquini L et al. showcased accuracy 83% accuracy for IDH mutation [33]. For MGMT methylation, the classification accuracies have ranged between 80.27 % - 95.12 % [20], [34]. Rathore et al. have demonstrated 88.28% accuracy for MGMT methylation marker using SVM [35]. For 1p/19q codeletion status, Akus Z. et al have illustrated 87.7% accuracy [36] whereas Yogananda C. et al have demonstrated 93.46% accuracy [37]. All these studies were carried out with only a single task whereas our multi-task model accounts for the similarity between tasks and utilizes the shared information, resulting in inductive learning that facilitates precise delineation of IDH, MGMT, 1p/19q and grade. Our model demonstrated an average test accuracy of 82.35% and the results were verified using the task wise and modality wise ablation analysis. In the ablation study of three task analysis, grade, IDH mutation and 1p/19q codeletion demonstrate a superior performance whereas in two task analysis IDH mutation and 1p/19q report the highest accuracies. Moreover, IDH mutation status shows the most significant performance in the single-task analysis and the result is at par with the previous studies. In addition to that the ablation analysis with T1CE and FLAIR being the most critical modalities for the tumor study, have demonstrated the highest accuracies in modality wise ablation analysis. Here we have modified the model to account for unlabeled data. Additionally, adding hierarchy in the model by keeping IDH wildtype and 1p/19q codeletion mutually exclusive helped the algorithm to learn the context between the features in possibly the most effective way. Overall, learning simultaneous multiple tasks has the potential to capture appropriate features that can consequently facilitate superior accuracies. Furthermore, the CNN model allows for better reproducibility and normalization within the layer and may potentially account for scanner variability.

To provide a better understanding and interpretability of the model, we have used high resolution class activation maps (HR-CAMs) that can localize the tumor region that largely contributes towards accurate prediction. From the maps, it can be interpreted that the model focuses on the most discerning areas of the tumoral region.
TABLE 4. Results of multitask network for each outcome task with and without autoencoders.

|                      | With Autoencoder |                      |                      |                      |                      |                      |
|----------------------|------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|                      | Precision        | F1-Score             | Sensitivity          | Specificity          | Training Accuracy    | Testing Accuracy     |
| Grade                |                  |                      |                      |                      |                      |                      |
| G3                   | 0.76             | 0.84                 | 0.64                 | 0.95                 | 82.51                | 79.68                |
| G4                   | 0.96             | 0.77                 | 0.64                 | 0.95                 |                      |                      |
| IDH Mutation         |                  |                      |                      |                      |                      |                      |
| MT                   | 0.77             | 0.8                  | 0.73                 | 0.84                 | 83.12                | 80.04                |
| WT                   | 0.84             | 0.78                 |                      |                      |                      |                      |
| 1p/19q Codeletion    |                  |                      |                      |                      |                      |                      |
| Codeleted            | 0.69             | 0.67                 | 0.98                 | 0.65                 | 93.04                | 88.73                |
| Non-Codeleted        | 0.92             | 0.95                 |                      |                      |                      |                      |
| MGMT Promoter Methylation |            |                      |                      |                      |                      |                      |
| Methylated           | 0.79             | 0.83                 | 0.66                 | 0.87                 | 84.34                | 80.96                |
| Unmethylated         | 0.81             | 0.73                 |                      |                      |                      |                      |

|                      | Without Autoencoder |                      |                      |                      |                      |                      |
|----------------------|---------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|                      | Precision           | F1-Score             | Sensitivity          | Specificity          | Training Accuracy    | Testing Accuracy     |
| Grade                |                     |                      |                      |                      |                      |                      |
| G3                   | 0.79               | 0.80                 | 0.67                 | 0.94                 | 83.57                | 73.81                |
| G4                   | 0.82               | 0.73                 |                      |                      |                      |                      |
| IDH Mutation         |                     |                      |                      |                      |                      |                      |
| MT                   | 0.65               | 0.74                 | 0.77                 | 0.81                 | 79.26                | 74.85                |
| WT                   | 0.76               | 0.68                 |                      |                      |                      |                      |
| 1p/19q Codeletion    |                     |                      |                      |                      |                      |                      |
| Codeleted            | 0.15               | 0.26                 | 0.45                 | 1.00                 | 94.58                | 89.55                |
| Non-Codeleted        | 1.00               | 0.51                 |                      |                      |                      |                      |
| MGMT Promoter Methylation |          |                      |                      |                      |                      |                      |
| Methylated           | 0.73               | 0.75                 | 0.46                 | 0.80                 | 70.33                | 61.15                |
| Unmethylated         | 0.56               | 0.51                 |                      |                      |                      |                      |

TABLE 5. Results of taskwise ablation analysis.

| Task                  | Accuracy      | Accuracy      | Accuracy      | Accuracy      | Accuracy      | Accuracy      |
|-----------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Grade                 | 79.68         | 82.65         | 81.78         | 76.85         | 84.68         | 83.16         |
| IDH Mutation          | 80.04         | 81.54         | 83.96         | 78.56         | 85.36         | 80.04         |
| 1p19q Codeletion      | 88.73         | 93.58         | 65.47         | 90.12         | 90.12         | 78.5          |
| MGMT Methylation      | 80.96         | 77.68         | 78.69         | 72.58         | 76.36         | 69.3          |

There are a few limitations to our study. Firstly, the study is implemented on a moderately sized dataset of 307 subjects. This is due to the constraints in the availability of multiple MRI sequences, histopathology and
immunohistochemistry record of all the patients. To address this drawback, we have augmented the data as deep learning algorithms work well with larger datasets. Secondly, we have included only T1-CE, T2 and FLAIR modalities of the MRI images and adding more modalities like ADC in future, could contribute significantly to the classification. Lastly, prospective testing is crucial before clinical translation in order to incorporate it effortlessly in the radiological workflow.

In summary, we employed a multitask learning approach, to identify critical genotypes like IDH, 1p/19q and MGMT simultaneously. By using semi-labeled and unlabeled data we alleviated the need for a priori expert information that is based on the invasive procedure and sophisticated lab work. In summary, the prediction of multiple molecular features concurrently is a step forward towards clinical translation of multi-modal image-based phenotypic glioma characterization that can not only help in patient-specific management but also may consequently support treatment efficacy and outcomes.

REFERENCES

[1] Q. T. Ostrom, L. Bauchet, F. G. Davis, I. Deltour, J. L. Fisher, C. E. Langer, M. Pekmezci, J. A. Schwartzbaum, M. C. Turner, K. M. Walsh, and M. R. Wrench, “The epidemiology of glioma in adults: A ‘state of the science’ review,” Neuro-Oncology, vol. 16, no. 7, pp. 896–913, 2014.

[2] A. Omuuro and L. DeAngelis, “Glioblastoma and other malignant gliomas: A clinical review,” Jama, vol. 310, no. 17, pp. 1842–1850, 2013.

[3] R. Jalali and A. Munshi, “Therapy for glioma: Indian perspective,” Indian J. Cancer, vol. 46, no. 2, p. 127, 2009.

[4] E. R. Laws, I. F. Parney, W. Huang, F. Anderson, A. M. Morris, A. Asher, K. O. Lillehei, M. Bernstein, H. Brem, A. Sloan, M. S. Berger, and S. Chang, “Survival following surgery and prognostic factors for recently diagnosed malignant glioma: Data from the glioma outcomes project,” J. Neurosurg., vol. 99, no. 3, pp. 467–473, Sep. 2003.

[5] D. N. Louis, A. Perry, G. Reifenberger, A. Von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, “The 2016 World Health Organization classification of tumors of the central nervous system: A summary,” Acta Neuropathol., vol. 131, no. 6, pp. 803–820, Jun. 2016.

[6] D. N. Louis, P. Wesseling, K. Aldape, D. J. Brat, D. Capper, I. A. Cree, and C. Eberhart, “CIMPACT-NOW update 6: New entity and diagnostic principle recommendations of the cIMPACT-Utrecht meeting on future CNS tumor classification and grading,” Brain Pathol., vol. 30, no. 4, pp. 844–856, Jul. 2020.

[7] H. Yan, D. W. Parsons, G. Jin, R. McLendon, B. A. Rasheed, W. Yuan, I. Kos, I. Batinic-Haberle, S. Jones, J. G. J. Riggins, and H. Friedman, “IDH1 and IDH2 mutations in gliomas,” New England J. Med., vol. 360, no. 8, pp. 765–773, 2009.

[8] C. Lu, P. S. Ward, G. S. Kapoor, D. Rohle, S. Turcan, O. Abdel-Wahab, C. R. Edwards, R. Khanin, M. E. Figueroa, A. Melnick, K. C. Wellen, D. M. O’Rourke, S. L. Berger, T. A. Chau, R. L. Levine, I. K. Mellinghoff, and C. B. Thompson, “IDH mutation impairs histone demethylation and results in a block to cell differentiation,” Nature, vol. 483, no. 7390, pp. 474–478, Mar. 2012.

[9] N. Hu, R. Richards, and R. Jensen, “Role of chromosomal 1p/19q co-deletion on the prognosis of oligodendrogliomas: A systematic review and meta-analysis,” Interdiscip. Neurosurg., vol. 5, pp. 58–63, 2016.

[10] S. H. E. Boots-Sprenger, A. Sijben, R. Jijntjes, B. B. J. Tops, A. J. J. Meena, A. L. Rivera, F. E. Bleeker, A. M. Gijtenbeek, K. Diefes, L. Heathcock, K. D. Aldape, J. W. M. Jeuken, and P. Wesseling, “Significance of complete 1p/19q co-deletion, IDH1 mutation and MGMT promoter methylation in gliomas: Use with caution,” Modern Pathol., vol. 26, no. 7, pp. 922–929, Jul. 2013.

[11] A. Idibai, Y. Marie, G. Pierroin, C. Brennetot, K. Hoang-Xuan, M. Kujas, K. Mokhtari, M. Sanson, J. Lejeune, A. Aurias, O. Delattre, and J.-Y. Delattre, “Two types of chromosome 1p losses with opposite significance in gliomas,” Ann. Neurol., vol. 58, no. 3, pp. 483–487, Sep. 2005.

[12] C. Daumas-Dupont, “Histological grading of gliomas,” Current Opinion Neurol. Neurosurg., vol. 5, no. 6, pp. 924–931, 1992.

[13] P. Kleihues, F. Solyemezoglou, B. Schäuble, B. Scheithauer, and F. C. Burger, “Histopathology, classification, and grading of gliomas,” Glia, vol. 15, no. 3, pp. 211–221, Nov. 1995.

[14] A. Sanz-Antoñanzas, J. Auzmendi-Iriarte, E. Carrasco-Garcia, L. Moreno-Cugnon, I. Ruiz, J. Villanua, L. Egaña, D. Otaegui, N. Samprón, and A. Matheu, “Liquid biopsy in glioblastoma: Opportunities, applications and challenges,” Cancers, vol. 11, no. 7, p. 950, Jul. 2019.

[15] S. Wu, J. Meng, Q. Yu, P. Li, and S. Fu, “Radiomics-based machine learning methods for isocitrate dehydrogenase genotype prediction of diffuse gliomas,” J. Cancer Res. Clin. Oncol., vol. 145, no. 3, pp. 543–550, Mar. 2019.

[16] H.-H. Cho and H. Park, “Classification of low-grade and high-grade glioma using multi-modal image radionics features,” in Proc. 39th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 3081–3084.
VANI SANTOSH is a Diagnostic Neuropathologist of repute, both nationally and internationally. She has been providing expert neuropathology services to facilitate patient care throughout our country and neighboring countries, for more than 25 years. She is currently working as a Professor of neuropathology at the National Institute of Mental Health and Neuroscience, Bengaluru. Her research focus centered around biomarker discovery in gliomas which resulted in three patents and over 100 publications in this field. She has authored 175 publications in indexed international and national journals and 19 chapters in various books, including the 2016 edition of the WHO Classification of Tumors of the Nervous System. Her research interest includes neuro-oncology, where she has made significant contributions towards brain tumor research by obtaining several competitive grants.

MADHURA INGALHALIKAR received the Ph.D. degree in biomedical engineering from the University of Iowa, Iowa City. She was a Postdoctoral Researcher and a Research Associate at the University of Pennsylvania, Philadelphia, PA, USA. She has had cutting-edge experience in the field of medical image analysis for the past 15 years. She has extensively worked in the area of neuroimaging with focus on diffusion MRI, multi-modality imaging, and deep learning. She is currently involved in multiple clinical studies in movement disorders, tumors as well as projects in cognitive science.

JITENDER SAINI is currently an Additional Professor and a practicing neuro-radiologist at the National Institute of Mental Health and Neuroscience, Bengaluru. He specializes in neuroimaging and has more than 14 years’ experience in clinical neuroimaging. He has a special interest in quantitative MR imaging especially in brain tumors. He has more than 150 articles in the peer-reviewed journals. He was a recipient of several grants.