



Machine Learning Models for Multiparametric Glioma Grading With Quantitative Result Interpretations

Xiuying Wang^{1*}, Dingqian Wang^{1†}, Zhigang Yao^{2†}, Bowen Xin¹, Bao Wang³, Chuanjin Lan³, Yejun Qin², Shangchen Xu⁴, Dazhong He⁵ and Yingchao Liu^{4*}

¹ School of Information Technologies, The University of Sydney, Sydney, NSW, Australia, ² Department of Pathology, Provincial Hospital Affiliated to Shandong University, Jinan, China, ³ School of Medicine, Shandong University, Jinan, China, ⁴ Department of Neurosurgery, Provincial Hospital Affiliated to Shandong University, Jinan, China, ⁵ School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China

OPEN ACCESS

Edited by:

Yun Zhou,
Washington University School
of Medicine in St. Louis, United States

Reviewed by:

Kuanquan Wang,
Harbin Institute of Technology, China
Jie-Han Liow,
National Institutes of Health (NIH),
United States
Shuyu Li,
Beihang University, China

*Correspondence:

Xiuying Wang
xiu.wang@sydney.edu.au
Yingchao Liu
13805311573@126.com

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 24 October 2018

Accepted: 24 December 2018

Published: 11 January 2019

Citation:

Wang X, Wang D, Yao Z, Xin B,
Wang B, Lan C, Qin Y, Xu S, He D
and Liu Y (2019) Machine Learning
Models for Multiparametric Glioma
Grading With Quantitative Result
Interpretations.
Front. Neurosci. 12:1046.
doi: 10.3389/fnins.2018.01046

Gliomas are the most common primary malignant brain tumors in adults. Accurate grading is crucial as therapeutic strategies are often disparate for different grades and may influence patient prognosis. This study aims to provide an automated glioma grading platform on the basis of machine learning models. In this paper, we investigate contributions of multi-parameters from multimodal data including imaging parameters or features from the Whole Slide images (WSI) and the proliferation marker Ki-67 for automated brain tumor grading. For each WSI, we extract both visual parameters such as morphology parameters and sub-visual parameters including first-order and second-order features. On the basis of machine learning models, our platform classifies gliomas into grades II, III, and IV. Furthermore, we quantitatively interpret and reveal the important parameters contributing to grading with the Local Interpretable Model-Agnostic Explanations (LIME) algorithm. The quantitative analysis and explanation may assist clinicians to better understand the disease and accordingly to choose optimal treatments for improving clinical outcomes. The performance of our grading model was evaluated with cross-validation, which randomly divided the patients into non-overlapping training and testing sets and repeatedly validated the model on the different testing sets. The primary results indicated that this modular platform approach achieved the highest grading accuracy of 0.90 ± 0.04 with support vector machine (SVM) algorithm, with grading accuracies of 0.91 ± 0.08 , 0.90 ± 0.08 , and 0.90 ± 0.07 for grade II, III, and IV gliomas, respectively.

Keywords: glioma grading, machine learning, morphological features, support vector machine, digital pathology images

INTRODUCTION

Gliomas are the most common primary malignant brain tumors in adults, accounting for 30% of all primary central nervous system (CNS) tumors and 80% of all malignant brain tumors (Goodenberger and Jenkins, 2012). According to their histological characteristics, such as cellularity, pleomorphism, nuclear atypia, necrosis, and endothelial proliferation, gliomas can be classified into WHO grades I–IV (Louis et al., 2007). Correctly differentiating tumor grades

is critical because they are widely used to predict patient outcomes and determine the use of adjuvant therapy protocols including aggressive radiotherapy and concomitant chemotherapy (Hsu et al., 1997). A precise prediction of the prognosis depends on an accurate pathology diagnosis. Compared to glioblastoma (GBM), low-grade gliomas (LGGs; WHO II) have greater chemotherapy sensitivity and a better post-therapy prognosis (Cillekens et al., 2000; Cavaliere et al., 2005; Wang and Jiang, 2013).

Considered as one of the “gold standards” and conventionally used by pathologists for tumor grading, haematoxylin and eosin (H&E) stained images provide specific histological characteristics and patterns to differentiate gliomas grades (Rubin et al., 2008). The parameters extracted by clinicians from WSIs mostly focused on the cellular level, including the number and size of nuclei, which reflects the proliferation and differentiation of the nuclei per unit area and heterogeneity of different levels of the tumor cells. For instance, compared to grade II gliomas, grade III, and grade IV both exhibit microvascular proliferation (MVP), indicating the presence of proliferation of enlarged blood vessels in the tissue (Burger et al., 2002). Grade IV gliomas can be further distinguished from grade III by examining H&E stained images for the presence of highly pleomorphic cells with hyperchromatic, irregular nuclei and brisk mitotic activity. Grade IV gliomas are often more mitotically active, necrosis prone, and generally associated with neovascularity, infiltration of surrounding tissue and a rapid postoperative progression (Louis et al., 2016). However, the conventional grading procedure based on H&E pathology images is often subjective and operator-dependent and of low reproducibility due to the inter-observer variability.

Apart from H&E stained images, immunocytochemical staining with Ki-67 antibodies has been widely accepted as an alternative reference for assessing the proliferative potential in tumor cells. Strictly associated with cell proliferation, Ki-67 nuclear antigen is present during all active phases of the cell cycle (Gap 1, Synthesis, and Gap 2 phases of the cell cycle) and mitosis but absent in resting (quiescent) cells (Gap 0) (Bruno and Darzynkiewicz, 1992). The proliferative index (PI), determined by Ki-67 immunohistochemistry (IHC), correlates well with the histological malignancy grade of gliomas (Skjulsvik et al., 2014). PI values may assist in differentiating grade II from III or grade II from IV; however, due to the overlapping PI values of grades III and IV, PI itself would not be considered as sufficient evidence to adequately determine these two malignancy grades (Skjulsvik et al., 2014).

Computer-assisted grading systems are capable of mining a large number of quantitative features from digital pathology slide images to sort out the most important patterns for grading. This ability provides the opportunity for better quantitative modeling of the disease appearance and hence possibly improves the prediction accuracy of tumor grading. In addition, it also provides a more reproducible, less labor-intensive and more efficient mechanism than manual grading by pathologists. However, the compelling opportunities offered by big digital pathology data come with optimal computational algorithm challenges (Madabhushi and Lee, 2016). For example, image analysis and computer-assisted detection models inadequately

address the data density in high-resolution digitized whole slide images (WSI). Mousavi et al. (2015) proposed to distinguish LGGs from GBMs using image features extracted from regions of interest (ROIs) in WSIs. However, the manual selection of ROI in this method may introduce a potential risk of inter-observer variance. Kong et al. (2009) tiled the WSI into non-overlapping pieces and analyzed each image tile to grade LGGs from GBMs. While fully processing all the image tiles, the method might be less computationally efficient, in particular, for high-resolution H&E stained images. To improve computational efficiency of the automated prognosis of neuroblastoma from H&E images, Sertel et al. (2009) proposed a multiresolution approach to extract local binary pattern and texture features of different scales.

While the above mentioned automated methods process and analyze WSIs for grading, the method proposed by Ertosun and Rubin (2015) classified LGGs and high-grade glioma (HGGs) on the basis of automated segmentation and analysis of cell nuclei and morphological features (Kong et al., 2013) but neglected the tumor patterns important for manual grading. To systematically compare the performance of classic machine learning models on grading, Huang and Lee (2009) implemented three machine learning models, Bayesian, k-nearest neighbors (KNN) and support vector machine (SVM), and concluded that KNN and SVM both achieved the highest accuracy. Apart from conventional machine learning, Reza and Iftekharuddin, 2016 proposed a deep learning-based grading system, characterizing each tile type with convolutional neural network (CNN). However, these machine learning and deep learning models are considered as “black boxes” without transparent interpretation of either the models themselves or the grading results.

To improve reproducibility and avoid inter-operator variance in conventional manual grading, in this study, we design and implement a platform for automated grading gliomas into grades II, III, and IV from digital pathology images. In our approach, discriminative visual, sub-visual and IHC parameters are identified, and a reliable machine learning model is selected. With the machine learning based models, we integrate information from both histological morphology images and proliferation biomarkers into a single unified framework to predict the glioma grade in 116 patients, which surpass the current clinical paradigm for patients diagnosed with glioma. Our platform provides interpretation on the grading outcomes to disclose the contributions of multiparametric features to individual cases and presents an alternative for objective, accurate, and interpretable prediction of glioma grading in the clinic.

MATERIALS AND METHODS

Material and Dataset

The dataset used in this study involves 146 cases of glioma grading from grades II to IV: 49 grade II, 45 grade III, and 52 grade IV images. All cases were from Shandong Provincial Hospital affiliated to Shandong University, and the pathology diagnoses of the cases were based on WHO standards (Louis et al., 2007). Paraffin-embedded samples were cut into 3 μm thick sections and stained with H&E stain. All H&E

images in this study were obtained from WSIs scanned by a Leica SCN400 slide scanner (Leica Biosystems, Nussloch, Germany) with multiresolution varying from 20× to 40×. Ki-67 immunohistochemical staining was prepared using an automated staining instrument (Ventana, Benchmark Ultra). As defined in Eq. 1, Ki-67 PI is the percentage of the number of immunoreactive tumor cells in relation to the total number of cells. At least 1000 tumor cells or alternatively, three high-power fields (HPFs) were examined by two independent experienced observers. The mean of Ki-67 PI is the average of the values calculated by different observers.

$$PI = \frac{\text{positive cells}}{\text{total cells}} * 100 \quad (1)$$

Automated Interpretable Glioma Grading With Machine Learning Models

As illustrated in **Figure 1**, our automated grading framework is composed of five major components including automated ROI identification, feature extraction, important feature selection, automated grading, and result interpretation.

Automated ROI Identification

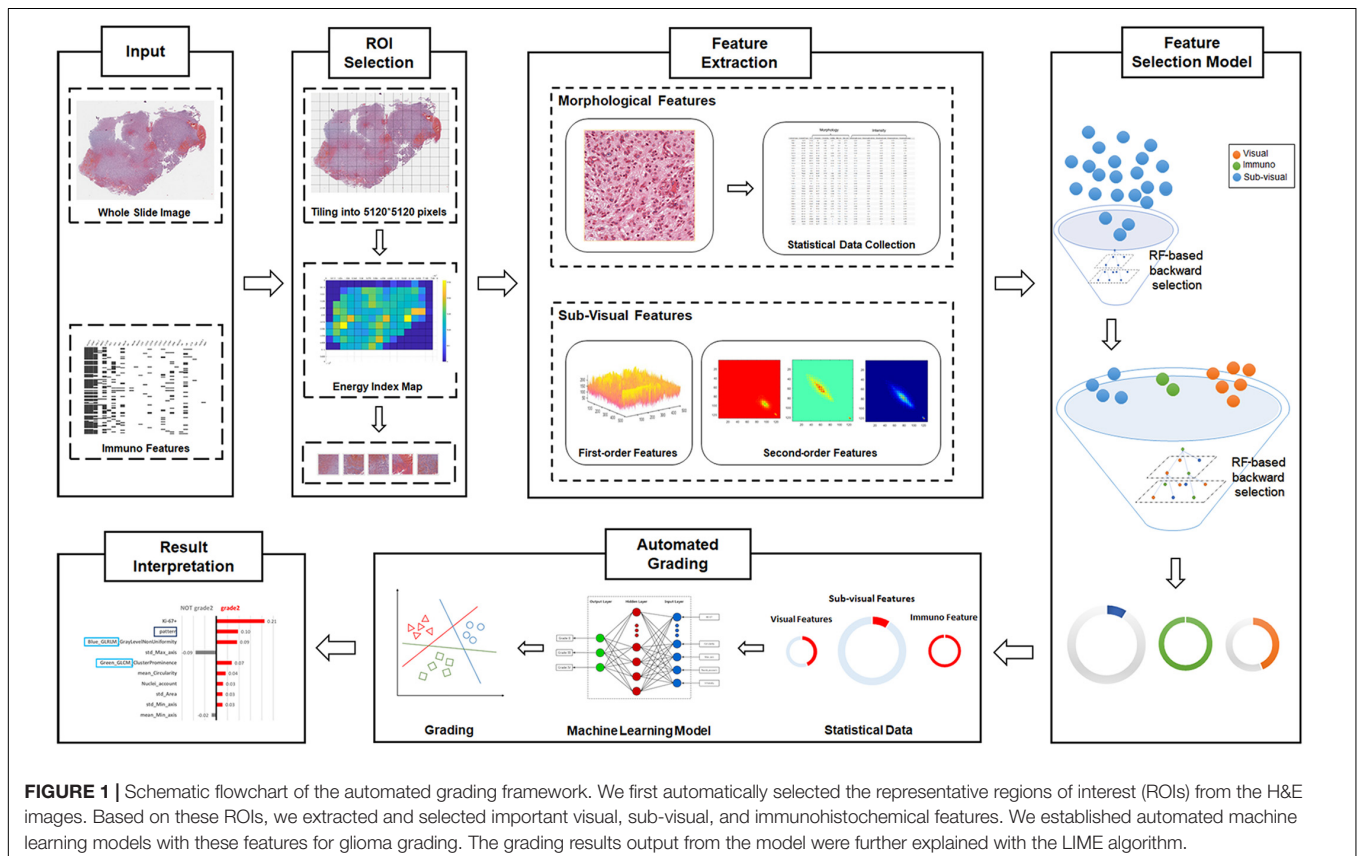
Our proposed computational method automatically identifies the ROIs that reflect cell proliferation and cellular density based on the number of nuclei in the regions. H&E images were first partitioned as tiles, each with a resolution of 5120*5120 pixels

to enable the process of high-resolution imaging (Sertel et al., 2009; Mobadersany et al., 2018). Then, we detected nuclei using watershed nuclei detection algorithm in each tile (Al-Kofahi et al., 2010; Kumar et al., 2017) and based on the density of the detected nuclei, the five tiles with the highest densities of nuclear were identified as the ROIs.

Multi-Parameter Extraction and Important Parameter Selection

From the identified ROIs, we extracted multi-parameters including visual features and sub-visual features as the inputs for automated grading. The visual features, including seven nuclear morphological features, five nuclear staining features, and nuclei clusters or patterns, were extracted to reflect a basis for observation when pathologists make diagnostic decisions (Kong et al., 2013). For instance, nuclear morphological characteristics such as shape, size, and circularity that reflect cellular atypia have been commonly used by pathologists to distinguish different grades of glioma.

In addition to the visual features, sub-visual features also contribute to accurate glioma grading. The sub-visual features are the computerized high-throughput intensity and texture image features that have been proved to have diagnostic, predictive, and prognostic power, although these features are somehow beyond human perception capabilities (Thawani et al., 2017). In our method, for instance, intensity features describe the first-order statistical information of the image intensity



distribution, while the second-order grey-level co-occurrence matrix (GLCM) features capture both statistical intensity and relationship between neighborhood pixels, revealing information such as homogeneity, contrast and entropy (Selvarajah and Kodituwakku, 2011). As illustrated in **Figure 1**, we used a random forest (RF)-based feature selection method combined with backward feature elimination to choose the most representative and informative features (Saeys et al., 2007).

Machine Learning Model for Automated Grading

To select the best machine learning model for grading, we compared the performance of four types of classic machine learning models including RF (Liaw and Wiener, 2002), gradient boosting decision tree (GBDT) (Friedman, 2002), SVM (Furey et al., 2000) and neural network (Hagan et al., 1996) (NN) (Zacharaki et al., 2009; Kothari et al., 2013). To achieve the best performance for each grading model, the hyperparameters of the model need to be optimized, for example, the tree number in RF model. The framework for automatically tuning the hyperparameters is provided in **Figure 2**.

Validation and Evaluation

116 patients out of 146 was used to train the model while the other 30 patients was used for further validate the model. The 116 patients were randomly divided into training datasets (75%) and testing datasets (25%) to test performance and validate the models (Larsen and Goutte, 1999). The corresponding training set was used to train the classifier, and the test set was used to verify the model performance. This procedure was repeated 30 times by changing random states while maintaining the same train-test split ratio. Apart from accuracy, other measurements including precision, recall, F1 score and confusion matrix were calculated to more completely validate grading performance. In addition, extra 30 cases were used as validation dataset with 30 times of cross-validation to further test the performance of our proposed model.

RESULTS

Multi-Parameter Extraction and Selection

In the multi-parameter extraction phase, we extracted 24 visual parameters and 171 sub-visual parameters. Since the

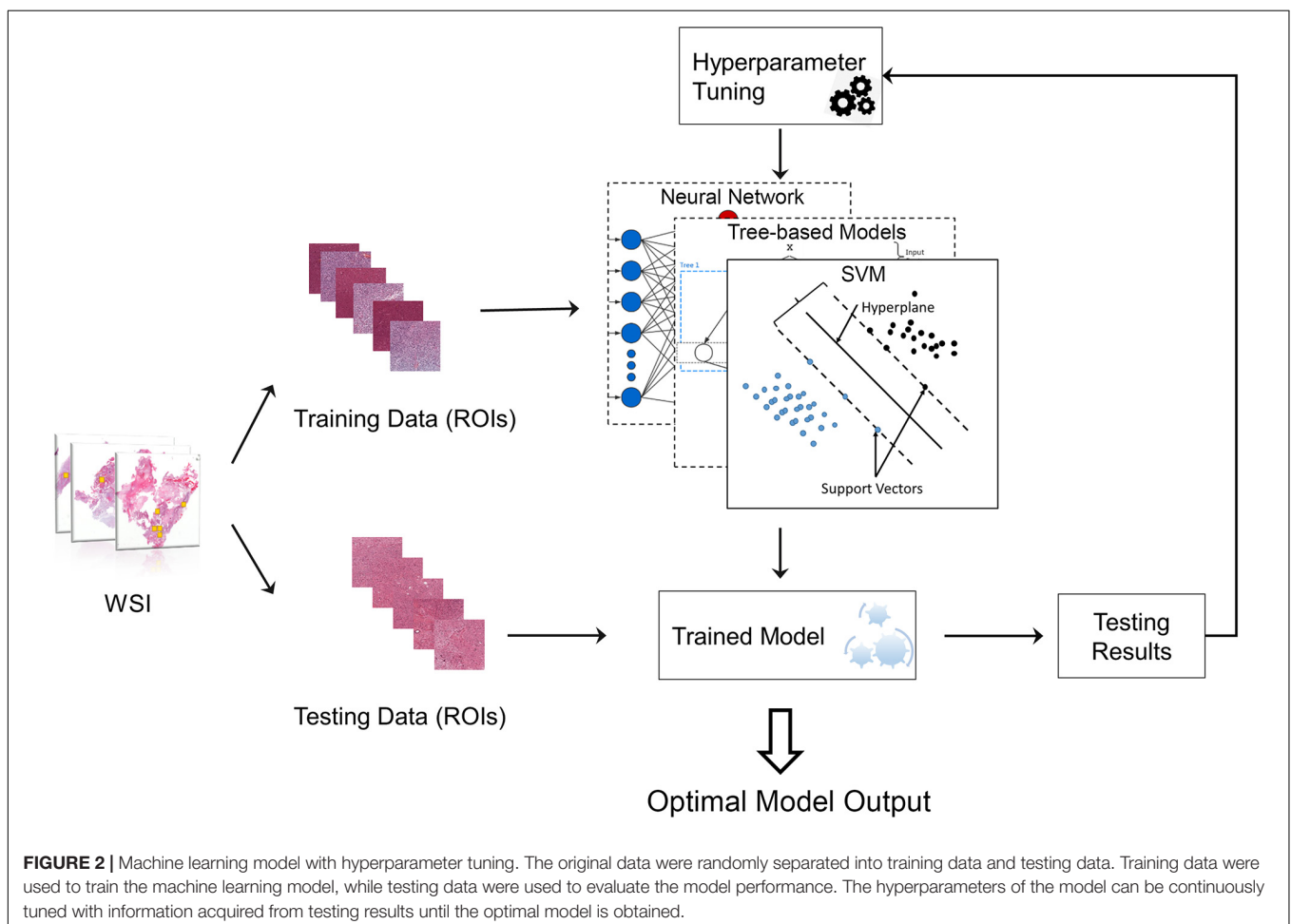


FIGURE 2 | Machine learning model with hyperparameter tuning. The original data were randomly separated into training data and testing data. Training data were used to train the machine learning model, while testing data were used to evaluate the model performance. The hyperparameters of the model can be continuously tuned with information acquired from testing results until the optimal model is obtained.

large size of the sub-visual features may suppress important morphological features, we performed the feature selection process on the sub-visual features twice to eventually choose 15 features from this category. The predictive performance of the selected 11 visual features, 15 sub-visual features and one IHC characteristic – Ki-67 in this study – were assessed with separate RF models. The box-plot in **Figure 3A** shows that visual features had the highest predictive power with 0.76 accuracy, while the accuracy for sub-visual features and Ki-67 only reached 0.62 and 0.53, respectively.

Automated Grading Performance With Different Models

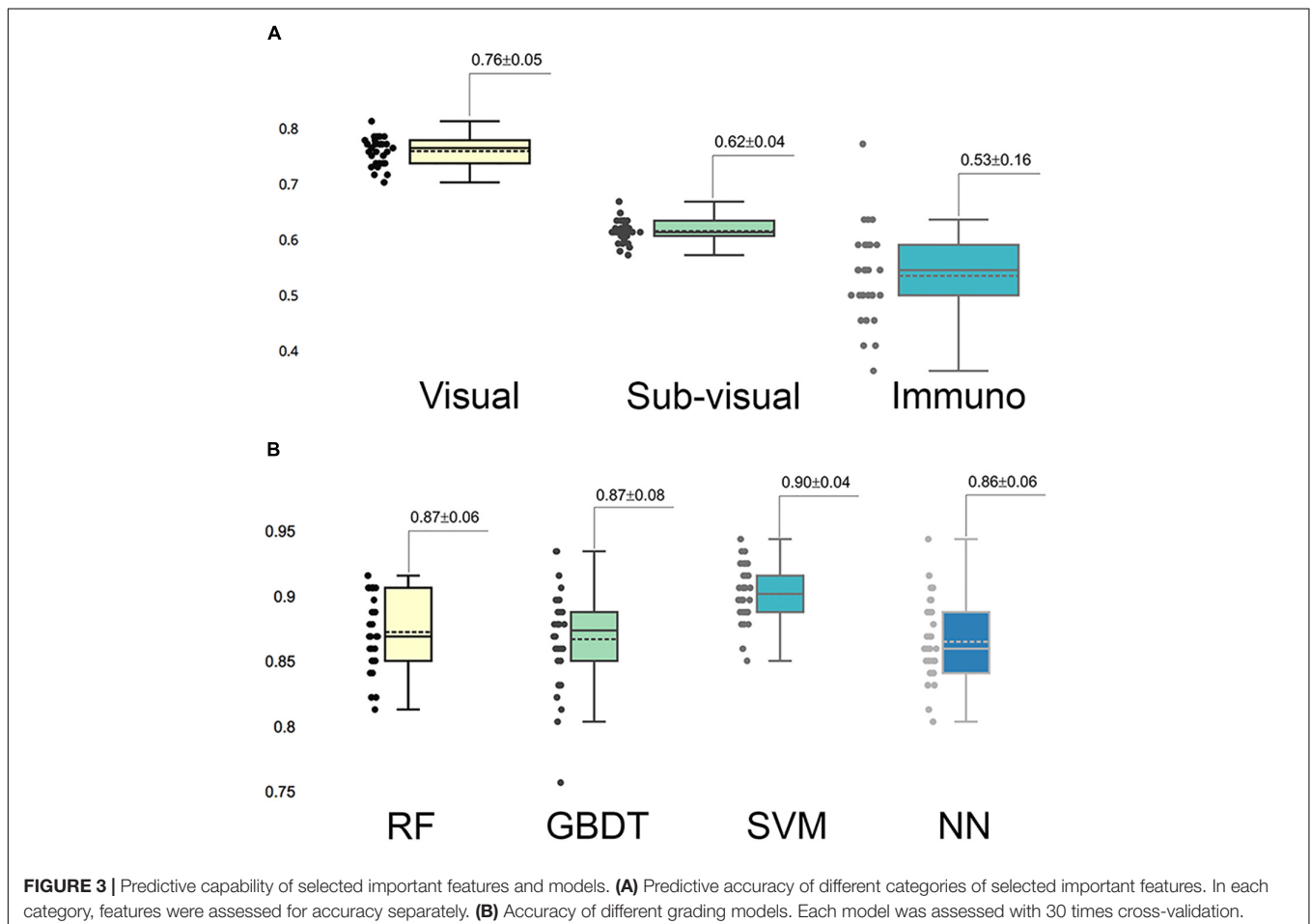
We fed these selected important parameters into the four machine learning models RF, GBDT, SVM, and NN to sort out the best grading model. Based on cross-validation, SVM achieved the most accurate and stable grading performance, with the accuracy of 0.90 ± 0.04 , while the other three models had lower but nearly the same performance with accuracy stabilized at approximately 0.87 (**Figure 3B**). **Table 1** shows the results of other measurements including F1, precision and recall, among which the SVM model maintained the best performance.

Further Investigation of SVM Performance

To better understand the performance of the SVM model for grading histologic grades of glioma II, III, and IV, we calculated a confusion matrix to uncover misclassification cases. As illustrated in **Figure 4A**, grade II had the highest accuracy (0.91) in the SVM model while the other two grades both had the accuracy of 0.90. As shown in the confusion matrix for the SVM model (**Figure 4B**), grade III gliomas may share common histological characteristics with both grade II and grade IV gliomas and thereby may lead to misclassification. In addition, we added 30 new patients of grade II to IV for the further validation. The validation performance achieved 0.88 ± 0.14 with 30 times cross-validation.

Quantitative Grading Results Interpretation

The machine learning algorithm is usually complex and considered as a “black box” without explicit interpretation of the learning process or the outputs. To provide a better quantitative interpretable explanation for the grading results, we used Local Interpretable Model-Agnostic Explanations (LIME) algorithm (Ribeiro et al., 2016) to reveal the



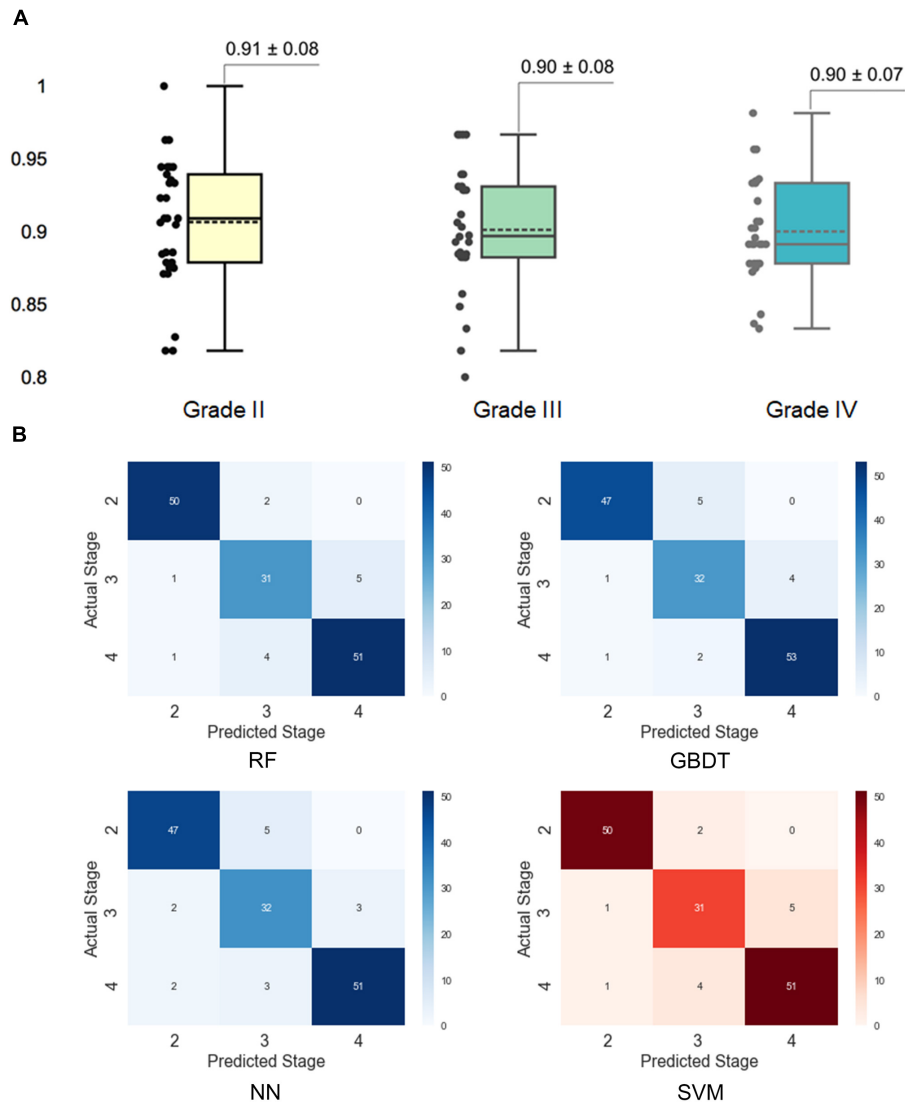


FIGURE 4 | Assessment of grading results with the support vector machine (SVM) model. **(A)** Prediction accuracy for different histological grades in the SVM model. The accuracy for each grade was obtained by separating the results by grades in the 30 times cross-validation. **(B)** Confusion matrixes for different models. The confusion matrixes reveal the number of misclassified cases for each grade.

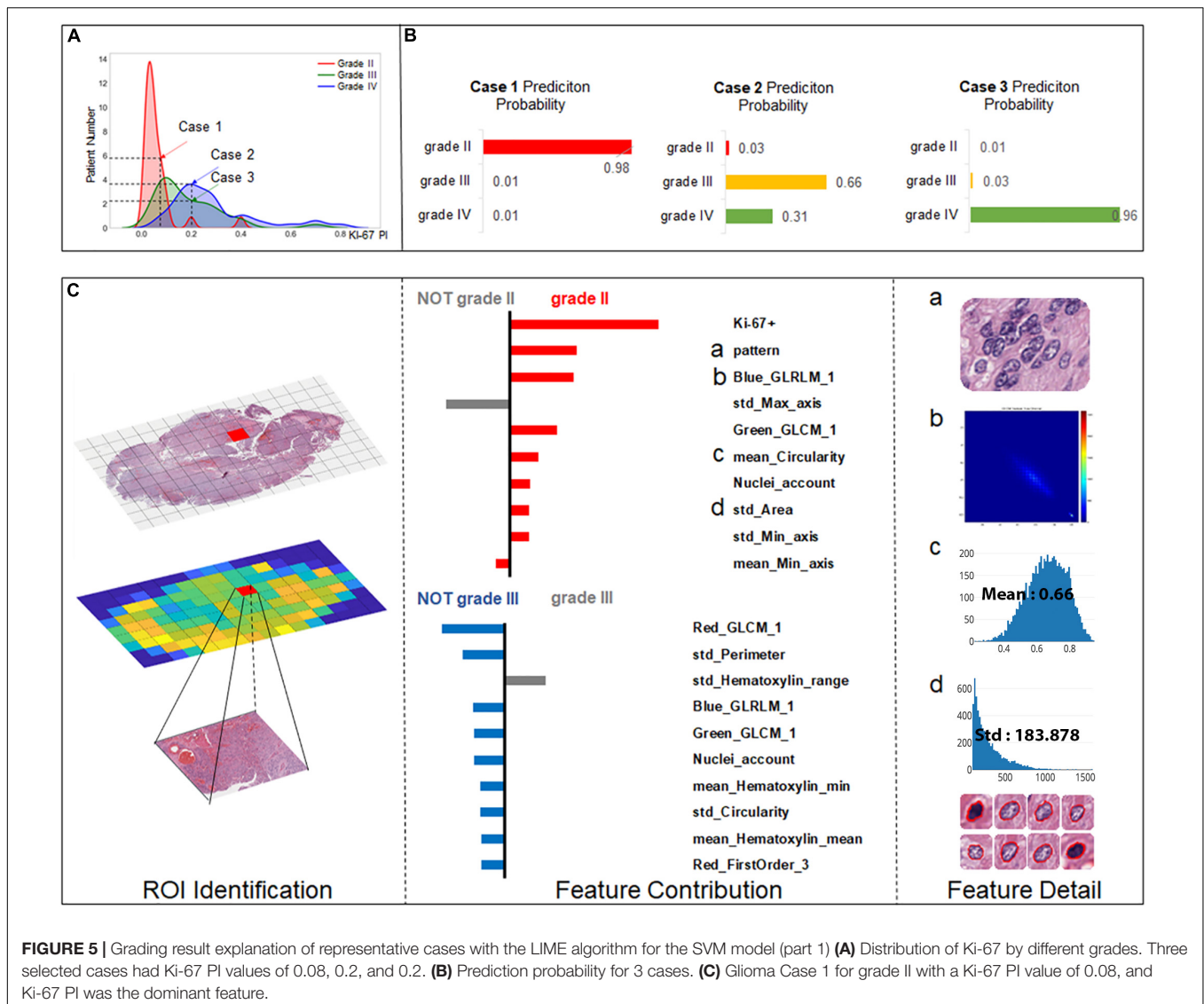
importance of features and their underlying contribution to the grading decision. In this section, we illustrated three case studies to interpret the automated grading results and the underlying reasoning for our SVM model with the LIME algorithm.

TABLE 1 | F1, accuracy, precision, and recall for different machine learning models.

Method	F1	Accuracy	Precision	Recall
Random forest	0.86 ± 0.07	0.87 ± 0.06	0.86 ± 0.07	0.87 ± 0.06
GBDT	0.86 ± 0.08	0.87 ± 0.07	0.86 ± 0.07	0.86 ± 0.07
Neural network	0.86 ± 0.05	0.86 ± 0.06	0.87 ± 0.06	0.87 ± 0.05
SVM	0.90 ± 0.07	0.90 ± 0.04	0.91 ± 0.04	0.91 ± 0.04

As Ki-67 has been proved to be a clinically significant indicator for grading, we predicted the grading results by machine learning models while also taking the influence of Ki-67 into consideration. To highlight this strategy, we chose three cases to elucidate the workflow of SVM for precise grading. As shown in **Figure 5A**, Case 1 is a relatively easy case to classify, being grade II with low Ki-67 (0.05). However, Case 2 and Case 3 are more ambiguous, both with a value of 0.2 for Ki-67, but one is grade III while the other is grade IV. We established a result explainer with LIME algorithm for the SVM model and then calculated the feature contribution for each case.

Case 1 (**Figure 5C**) was correctly classified as grade II with a high probability. As Ki-67 has been clinically proved to have the discriminative capacity to distinguish between LGG and HGG, it is not surprising that Ki-67 has the dominant influence



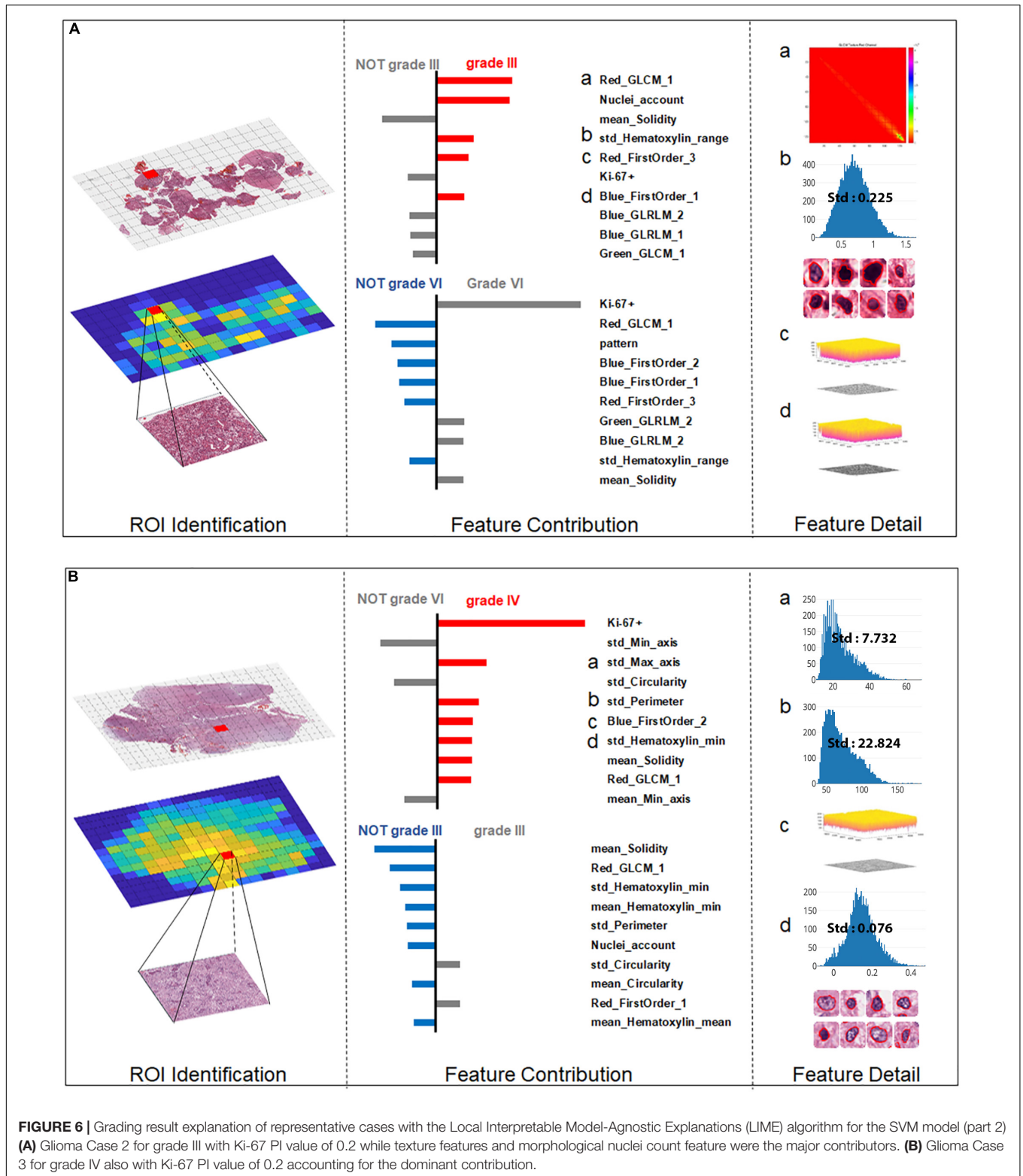
underlying this correct decision. Apart from Ki-67, visual features such as cell pattern and sub-visual features such as GLCM texture features also can correctly guide the grading. The result also shows that although some morphological features such as perimeter do not support the decision of grade II, they excluded grade III, which also contributed to accurate grading.

For the more ambiguous Case 2 (Figure 6A) and Case 3 (Figure 6B), with regard to Ki-67 itself, our classification model demonstrated its capability for an accurate decision. Case 2 (Figure 5B) was the most challenging case among these three cases because in this case, Ki-67 had only a negative influence on the grading decision; however, with the strong support of GLCM texture features and the morphological nuclei count feature, our grading system made a correct decision. In Case 3, Ki-67 was the major contributor to the correct decision, while the morphological features such as standard deviation of cells' max axis and standard deviation of cells' perimeter were the second and third contributors.

DISCUSSION

Given the well-recognized difficulties associated with the grading of gliomas, in this study, we developed an automated and interpretable grading of gliomas with a machine learning platform to provide an additional reference for diagnostic decision making and patient management and prognosis, which is more efficient, effective, and objective than currently used methods. Our major findings were that (1) our method achieved a high grading accuracy of 90% for classifying gliomas into grades II, III, and IV; and (2) more importantly, for the first time, an interpretable insight into the grading outputs was provided for the histology grading system with the “black-box” machine learning models.

Computerized image perception in tissue histology is much more difficult due to the complexity of the cell patterns and inherent structural associations between different tissue components. Accurate histopathological diagnosis determines



patient management, treatment, and follow-up, so any method that results in more objective glioma grading may be of great value. In the current literature, automated grading either was based on texture patterns extracted from the selected ROIs

(Mousavi et al., 2015) to capture the histological structures in the histological images or utilized morphological features extracted from segmented tumor cells (Sertel et al., 2009; Ertosun and Rubin, 2015). Our method achieved high accuracy

through fully utilizing the selected informative visual and sub-visual features from H&E images combined with the Ki-67 PI value. To take into account the visual cues used by pathologists for manual grading, visual features such as texture patterns and cell morphological characteristics were extracted from selected ROIs. The sub-visual features that are beyond human perception capacities can help the precise grading through more comprehensive quantitative analyses. Furthermore, we innovatively integrated the IHC indicator Ki-67 PI into our grading platform, which contributed to boosting our grading accuracy with high discriminatory power.

Generally, the aim of this research was to objectify, standardize and quantify features that are already widely accepted as important by pathologists. Inter-observer discrepancies, which had long been a diagnostic problem in the past, could be partly overcome by employing the morphometric analysis used in the present study (Kinjo et al., 2008). The ability of the system to objectively identify regions of tumor may additionally complement the pathologist's diagnosis and assist in tailored treatment. In this study, we attempted to automatically clarify the cellular composition and histopathological features of different grades of gliomas by utilizing morphometric- and immunohistochemical-based machine learning models. We assessed the importance of visual features, sub-visual features and Ki-67 separately, and the highest predictive power was achieved by visual features, achieving an accuracy of 0.76, which indicates that the histomorphology of glioma, such as nuclear morphological features and nuclear staining features and patterns play an important role in glioma grading.

The interpretability of the individuals' grading results provided quantitative insights into the feature contributions of each individual case. Most of the grading systems based on machine learning models are considered "black boxes," and it would be valuable for patient management if clinically trusted reasoning could be revealed. In our study, we used the LIME algorithm to provide an explanation for each individual grading result. Case 1 and Case 3 in **Figures 5C, 6B** endorse the important role that Ki-67 plays in grading decisions, which is in accordance with the clinical fact that Ki-67 PI is positively correlated with the histopathology grades (Faria et al., 2006). However, Ki-67 sometimes has a detrimental influence on the grading process, as shown in Case 2, because Ki-67 PI of grade III gliomas largely overlaps with that of grade IV (Shibata et al., 1988; Kayaselçuk et al., 2002; Johannessen and Torp, 2006; Skjulsvik et al., 2014). As a result, Ki-67 itself cannot be the only determinant of grading. It has been accepted that the Ki-67 PI value is $3.0 \pm 2.1\%$ in grade II gliomas, $11.8 \pm 3.4\%$ in grade III anaplastic gliomas and $15.8 \pm 7.4\%$ in grade IV GBMs (Johannessen and Torp, 2006). Furthermore, the Ki-67 PI values of grade III anaplastic glioma can overlap with values of grade II glioma at one end of the range and with those of GBM at the other in the diagnostic practice of pathologists. Apart from Ki-67, morphological features including the standard deviation of cells' max axis and perimeter were also found to be a significant contributor to glioma grading because HGG often exhibits strong heterogeneity with irregular cell shapes (Shuangshoti et al., 2000; Louis et al., 2016). The precise morphological features extracted

from the most aggressive regions help to investigate intratumoral histological heterogeneity for precise histopathology grading (Walker et al., 2001). The primary grading results still support the conclusion that the only sure way to determine the histopathological WHO grade remains the pathohistological evaluation of the H&E stained tumor sample (Stoyanov et al., 2017).

With three representative cases we could gain a deeper understanding of why misclassified grade III cases are often labeled as grade II, while grade IV cases could be wrongly labeled grade III as shown in the confusion matrix. Actually, the Ki-67 PI values of the GBM group can be as low as those for grade II tumors, indicating the limitation of Ki-67 values in the overlap region. In our work, we introduced a number of histological sub-visual features such as the intensity and GLCM texture features to distinguish these three grades of gliomas. These insights become easy once it is understood what the algorithm models are actually doing, which in turn leads to models that generalize much better results. In many applications of machine learning, users are asked to trust a model to help them make decisions. There has always been a focus on "trust" in any type of modeling methodology, but with machine learning, many people feel that the black-box approach taken with these methods is not trustworthy. Through this machine learning-based approach, we could use the LIME algorithm to explain individual predictions to the decision-maker (the pathologist), and that understanding of the model's predictions can be an additional useful tool when deciding whether a model is trustworthy or not for the final diagnosis from a pathologist.

Our study did, however, have its limitations. First, grade I gliomas are not included in our studies. They are accurately considered benign in clinical practice, in that complete surgical excision is considered curative. Therefore, grade I gliomas are different from grade II-IV gliomas in biological behaviour. The results of our discrimination of grade III and grade IV are just reasonable preliminary results but leave much room for improvement. Considering that necrosis is one of the remarkable features of GBM, we plan to use cell necrosis as an input feature to further train models to distinguish grade III from grade IV GBM. Actually, morphometric data research has indicated that the cellularity of oligodendrogliomas type II was significantly higher than that of diffuse astrocytomas and that the conditional entropy of oligodendrogliomas type III was significantly lower than that of diffuse astrocytomas (Faria et al., 2006), so further stratification of LGG (grade II) will lead to tailored glioma management according to their different biological behaviour. Hence, future work will focus on improvements utilizing larger datasets, including multi-centre cases. In addition, there is still much room to improve the grading performance and generalization of the algorithm model.

CONCLUSION

In conclusion, our approach provides an objective alternative for quantitative pathology research and for the implementation

of morphological data in routine diagnostic practice. The machine learning model utilized multi-parameters including morphometric and sub-visual parameters as well as Ki-67 PI information to ensure high accuracy, efficiency and consistency in glioma grading. Interpretable grading platform has the potential to facilitate personalized medicine in the setting of malignant gliomas. With different important features identified for different patients, specific phenotypic tumor characteristics can be uncovered for optimal treatment selection. In addition, as our method is fully automated and quantitative with high reliability, it becomes easier for our platform to be introduced in routine clinical practice.

ETHICS STATEMENT

Does the study presented in the manuscript involve human or animal subjects: Yes.

This study was carried out in accordance with the recommendations of 1964 Declaration of Helsinki and its later amendments. Written Informed consent was obtained from all patients, and this study was approved by the Ethics Committee

REFERENCES

- Al-Kofahi, Y., Lassoued, W., Lee, W., and Roysam, B. (2010). Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Trans. Biomed. Eng.* 57, 841–852. doi: 10.1109/TBME.2009.2035102
- Bruno, S., and Darzynkiewicz, Z. (1992). Cell cycle dependent expression and stability of the nuclear protein detected by Ki-67 antibody in HL-60 cells. *Cell Prolif.* 25, 31–40. doi: 10.1111/j.1365-2184.1992.tb01435.x
- Burger, P. C., Scheithauer, B. W., and Vogel, F. S. (2002). *Surgical Pathology of the Nervous System and Its Coverings*. Oxford: Gulf Professional Publishing.
- Cavaliere, R., Lopes, M. B. S., and Schiff, D. (2005). Low-grade gliomas: an update on pathology and therapy. *Lancet Neurol.* 4, 760–770. doi: 10.1016/S1474-4422(05)70222-2
- Cillekens, J., Beliën, J., Van der Valk, P., Faes, T. J., van Diest, P. J., Broeckaert, M. A., et al. (2000). A histopathological contribution to supratentorial glioma grading, definition of mixed gliomas and recognition of low grade glioma with Rosenthal fibers. *J. Neurooncol.* 46, 23–43. doi: 10.1023/A:1006496328729
- Ertosun, M. G., and Rubin, D. L. (eds). (2015). Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks. *AMIA Annu. Symp. Proc.* 2015, 1899–1908.
- Faria, M. H., Gonçalves, B. P., do, Patrocínio RM, de, Moraes-Filho MO, and Rabenhorst, S. H. (2006). Expression of Ki-67, Topoisomerase II α and c-MYC in astrocytic tumors: correlation with the histopathological grade and proliferative status. *Neuropathology* 26, 519–527. doi: 10.1111/j.1440-1789.2006.00724.x
- Friedman, J. H. (2002). Stochastic gradient boosting. *Comput. Stat. Data Anal.* 38, 367–378. doi: 10.1016/S0167-9473(01)00065-2
- Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., and Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 16, 906–914. doi: 10.1093/bioinformatics/16.10.906
- Goodenberger, M. L., and Jenkins, R. B. (2012). Genetics of adult glioma. *Cancer Genet.* 205, 613–621. doi: 10.1016/j.cancergen.2012.10.009
- Hagan, M. T., Demuth, H. B., and Beale, M. H. (1996). *Neural Network Design*. Boston, MA: Pws Pub.
- Hsu, D. W., Louis, D. N., Efrid, J. T., and Hedley-Whyte, E. T. (1997). Use of MIB-1 (Ki-67) immunoreactivity in differentiating grade II and grade III gliomas. *J. Neuropathol. Exp. Neurol.* 56, 857–865. doi: 10.1097/00005072-199708000-00003

of Provincial Hospital affiliated to Shandong University (LCYJ: No2017-035).

AUTHOR CONTRIBUTIONS

XW, YL, and SX designed the experiments. YL, CL, and YQ conducted H&E slide imaging. ZY and CL performed the immunohistochemical analysis. XW, DW, BW, BX, and DH designed and implemented the computational platform and statistical analyses. XW, DW, BX, ZY, and YL drafted the manuscript. XW and YL initiated and supervised the entire project.

FUNDING

We gratefully acknowledge the financial support of the 2016 National Key Research and Development Program (2016YFC0106105) and the Taishan Scholars Program (No. tsqn20161070).

- Huang, P.-W., and Lee, C.-H. (2009). Automatic classification for pathological prostate images based on fractal analysis. *IEEE Trans. Med. Imaging* 28, 1037–1050. doi: 10.1109/TMI.2009.2012704
- Johannessen, A. L., and Torp, S. H. (2006). The clinical value of Ki-67/MIB-1 labeling index in human astrocytomas. *Pathol. Oncol. Res.* 12, 143–147. doi: 10.1007/BF02893360
- Kayaselçuk, F., Zorludemir, S., Gümürdühü, D., Zeren, H., and Erman, T. (2002). PCNA and Ki-67 in central nervous system tumors: correlation with the histological type and grade. *J. Neurooncol.* 57, 115–121. doi: 10.1023/A:1015739130208
- Kinjo, S., Hirato, J., and Nakazato, Y. (2008). Low grade diffuse gliomas: shared cellular composition and morphometric differences. *Neuropathology* 28, 455–465. doi: 10.1111/j.1440-1789.2008.00897.x
- Kong, J., Cooper, L. A., Wang, F., Gao, J., Teodoro, G., Scarpace, L., et al. (2013). Machine-based morphologic analysis of glioblastoma using whole-slide pathology images uncovers clinically relevant molecular correlates. *PLoS One* 8:e81049. doi: 10.1371/journal.pone.0081049
- Kong, J., Sertel, O., Shimada, H., Boyer, K. L., Saltz, J. H., and Gurcan, M. N. (2009). Computer-aided evaluation of neuroblastoma on whole-slide histology images: classifying grade of neuroblastic differentiation. *Pattern Recogn.* 42, 1080–1092. doi: 10.1016/j.patcog.2008.10.035
- Kothari, S., Phan, J. H., Stokes, T. H., and Wang, M. D. (2013). Pathology imaging informatics for quantitative analysis of whole-slide images. *J. Am. Med. Inform. Assoc.* 20, 1099–1108. doi: 10.1136/amiainf-2012-001540
- Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., and Sethi, A. (2017). A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans. Med. Imaging* 36, 1550–1560. doi: 10.1109/TMI.2017.2677499
- Larsen, J., and Goutte, C. (eds). (1999). “On optimal data split for generalization estimation and model selection,” in *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing IX* (Madison, WI: IEEE). doi: 10.1109/NNSP.1999.788141
- Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22.
- Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvett, A., et al. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* 114, 97–109. doi: 10.1007/s00401-007-0243-4
- Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization

- classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 131, 803–820. doi: 10.1007/s00401-016-1545-1
- Madabhushi, A., and Lee, G. (2016). *Image Analysis and Machine Learning in Digital Pathology: Challenges and Opportunities*. New York, NY: Elsevier.
- Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D. A., Barnholtz-Sloan, J. S., Velázquez Vega, J. E., et al. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci. U.S.A.* 115, E2970–E2979. doi: 10.1073/pnas.1717139115
- Mousavi, H. S., Monga, V., Rao, G., and Rao, A. U. (2015). Automated discrimination of lower and higher grade gliomas based on histopathological image analysis. *J. Pathol. Inform.* 6:15. doi: 10.4103/2153-3539.153914
- Reza, S. M., and Iftekharuddin, K. M. (eds). (2016). “Glioma grading using cell nuclei morphologic features in digital pathology images,” in *Proceedings of the SPIE Medical Imaging, Medical Imaging 2016*, Orlando, FL.
- Ribeiro, M. T., Singh, S., and Guestrin, C. (eds). (2016). “Why should i trust you?: Explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: ACM), 1135–1144. doi: 10.1145/2939672.2939778
- Rubin, R., Strayer, D. S., and Rubin, E. (2008). *Rubin’s Pathology: Clinicopathologic Foundations of Medicine*. Philadelphia, PA: Lippincott Williams & Wilkins.
- Saeyns, Y., Inza, I., and Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics* 23, 2507–2517. doi: 10.1093/bioinformatics/btm344
- Selvarajah, S., and Kodituwakku, S. (2011). Analysis and comparison of texture features for content based image retrieval. *Int. J. Latest Trends Comput.* 2, 108–113.
- Sertel, O., Kong, J., Shimada, H., Catalyurek, U. V., Saltz, J. H., and Gurcan, M. N. (2009). Computer-aided prognosis of neuroblastoma on whole-slide images: classification of stromal development. *Pattern Recogn.* 42, 1093–1103. doi: 10.1016/j.patcog.2008.08.027
- Shibata, T., Burger, P., and Kleihues, P. (eds). (1988). “Ki-67 immunoperoxidase stain as marker for the histological grading of nervous system tumours,” in *Proceedings of the 8th European Congress of Neurosurgery* (Barcelona: Springer), 103–106. doi: 10.1007/978-3-7091-8978-8_22
- Shuangshoti, S., Navalitloha, Y., Kasantikul, V., Shuangshoti, S., and Mutirangura, A. (2000). Genetic heterogeneity and progression in different areas within high-grade diffuse astrocytoma. *Oncol. Rep.* 7, 113–120. doi: 10.3892/or.7.1.113
- Skjulsvik, A. J., Mørk, J. N., Torp, M. O., and Torp, S. H. (2014). Ki-67/MIB-1 immunostaining in a cohort of human gliomas. *Int. J. Clin. Exp. Pathol.* 7, 8905–8910.
- Stoyanov, G. S., Dzhakov, D. L., Kitanova, M., Donev, I. S., and Ghenev, P. (2017). Correlation between Ki-67 index, World Health Organization grade and patient survival in glial tumors with astrocytic differentiation. *Cureus* 9:e1396. doi: 10.7759/cureus.1396
- Thawani, R., McLane, M., Beig, N., Ghose, S., Prasanna, P., Velcheti, V., et al. (2017). Radiomics and radiogenomics in lung cancer: a review for the clinician. *Lung Cancer* 115, 34–41. doi: 10.1016/j.lungcan.2017.10.015
- Walker, C., Joyce, K. A., Thompson-Hehir, J., Davies, M. P., Gibbs, F. E., Halliwell, N., et al. (2001). Characterisation of molecular alterations in microdissected archival gliomas. *Acta Neuropathol.* 101, 321–333.
- Wang, Y., and Jiang, T. (2013). Understanding high grade glioma: molecular mechanism, therapy and comprehensive management. *Cancer Lett.* 331, 139–146. doi: 10.1016/j.canlet.2012.12.024
- Zacharakis, E. I., Wang, S., Chawla, S., Soo, Yoo D, Wolf, R., Melhem, E. R., et al. (2009). Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn. Reson. Med.* 62, 1609–1618. doi: 10.1002/mrm.22147

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wang, Wang, Yao, Xin, Wang, Lan, Qin, Xu, He and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.