# Glioma Histopathological Images Classification with Deep CNN and Object Level Features

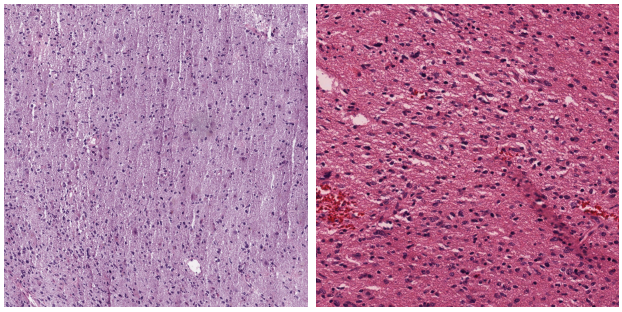
Daisuke Saito[1], Hiroharu Kawanaka[1], Shinji Tsuruoka[1] Bruce J. Aronow[2, 3, 4], V. B. Surya Prasath[2, 3, 4, 5, *]
[1] *Department of Electrical and Electronic Engineering, Mie University, Japan*
[2] *Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, OH 45229 USA*
[3] *Department of Pediatrics, University of Cincinnati, OH USA*
[4] *Department of Biomedical Informatics, College of Medicine, University of Cincinnati, OH 45267 USA*
[5] *Department of Electrical Engineering and Computer Science, University of Cincinnati, OH 45221 USA*
[*] *Email: surya.prasath@cchmc.org*

***Abstract*—In the field of histopathology, there are enormous amount of imaging samples available for analysis. Evaluating the various disease stages of brain tumors are an important task in oncology and requires the pathologist to analyze manually. However, the computational image processing techniques can help alleviate this burden and can also provide results quantitatively. Recently, the availability of large scale histopathology datasets coupled with advances in machine learning models provide robust tools for identifying disease stage classification in brian glioma automatically. Despite the improved results obtained with recent state-of-the-art deep convolutional neural network (CNN) classification models, how the models learned and the explainability is an important requirement in a trusted devising computer aided diagnosis (CAD) system. In this work, we discuss the classification reasoning within the context brain glioma histopathological disease stages. Moreover, we identify relationships between the decisions made by an adapted CNN classification model and object level features extracted using cell nuclei regions that are advocated in traditional computational pathology area. We further test an approach based on the regions of interest extracted by CNN along with object level features with support vector machine (SVM) to check the significance of each object level feature in the disease stages classification by our improved CNN model. Our experimental results indicate that our deep CNN model correctly classifies low versus high grade gliomas and with very low error rates on the cancer genome atlas histopathological images.**

***Index Terms*—Histopathological images, Brain glioma, deep learning, convolutional neural network, disease stage classification, explainable AI.**

## I. Introduction

In the field of oncology, digital pathology images play a vital role. Due to the availability big data whole slide imagery, there is a need to automatically evaluate samples for analysis. In the traditional setup, pathologists analyzes these images either completely manually or with some simple image processing tools, thus making the processing not efficient and cumbersome. Automatic machine learning models that can leverage the power of available data can help oncologists to evaluate disease stages and can help find the disease progression of patients. Recently, deep learning and especially convolutional neural networks (CNNs) are proven to provide highly accurate solutions to various medical image classification problems [1]. In computational pathology, such artificial intelligence (AI) solutions are entering mainstream disease stage classification.



(a) GBM (b) LGG

Fig. 1. An example of Histopathological images of Glioma from The Cancer Genome Atlas (TCGA). (a) Glioblastoma multiforme (GBM), (b) lower-grade Glioma (LGG). Note the distinct distribution of nuclei between low grade versus high grade. The shape, distribution, and morphology of the nuclei features can be used to predict the disease stage classification.

From such a background, there are many studies that are to analyze medical images with machine learning technology for diagnosis of each disease [2][3][4]. In our work, we focus on the pathological field of brain glioma. Brain glioma is one of the severe tumors with very poor prognosis for the affected patients. Glioma is one of the severe brain tumors. At previous works, Fukuma et al [5][6][7] have proposed a disease stage classification method that focuses on the relationship between the size and position of the cell nuclei region for automatically and quantitative analysis based on brain histopathological images. In addition, Yonekura et al [8] have proposed an approach using a modified CNN model to consider the features exclusive of the cell nuclei region. As well as, also Mehmet Günhan Ertosun et al have classified the glioma disease stage with CNN [9]. CNN is one of the machine learning and constructed by mainly the convolutional layer and pooling layer. And it extracts features for classification automatically. Using an adapted CNN model, they obtained a high classification accuracy with 96% for two disease stages of glioma, namely low versus high grades.

Despite the success in automatically classifying disease

stages, the CNN model cannot explain how the classification into each stages happened since feature selection is not apparent. Thus, we require visualizations of deep CNN models to be able to use the disease stage classification in a clinical setting. Moreover, certain object level features can still be of use and a combination with deep CNN model can provide explainable AI solutions for such challenging image classification problem. In this work, we discuss that explaining how CNN decides to classify into each disease stage for glioma. In particular, we focused on the relationship between the object level feature of the cell nuclei region and CNN decision. Our experimental results on the cancer genome atlas (TCGA) glioma histopathological images indicate that deep CNN models correctly classify high versus low grade gliomas reliably.

Rest of the work is organized as follows. Section II provides the details of our DL CNN model including the description of the object level features and visualizations. Section III provides our results on the TCGA database. Finally, Section IV concludes the paper.

## II. Deep CNN and Object Level Features for Classification

### A. Glioma Pathological Images

We focused on glioma pathological images stained with H & E staining. Glioma is one of brain tumor. Glioma is sorted 4 grades by the disease stages. In the case of grad 4 (Glioblastoma : GBM), a life expectancy is much shorter than the case of grade 3, 2 ( Low Grade Glioma : LGG). In this paper, glioma disease stages are regarded as 2 classes (GBM and LGG). Figure 1 shows examples of glioma pathological images at each stage.

### B. Data Set

Glioma pathological images are available at The Cancer Genome Atlas (TCGA - https://cancergenome.nih.gov). We obtained glioma pathological images from TCGA for experimental materials.

The amount of the image size obtained from TCGA is too large for the experiments using CNN. Then we patched the pathological images into $1000 \times 1000$ to evaluate glioma disease stages with CNN. And if the patched image is contained a cell region enough, we added it to the data set. Otherwise, the image were removed. Thus our data set has 20000 images (GBM : 10000, LGG : 10000).

In the next step, the experimental images were divided into a training data set (80%), validation data set (10%), test data set (10%). Also, in order to confirm the experiment accuracy using CNN, we made two subsets that were shuffled from the above data set with the same distribution.

### C. Evaluation of the Glioma Disease Stage with CNN

In this experiment, we applied CNN to the classification of each glioma disease stage. CNN is one of the most useful Deep Learning methods for the classification in image processing. Also, CNN is constructed mainly convolutional layer and max pooling layer. In particular, VGG16 has applied it with fine-tuning. Figure 2 shows our learning model. The original VGG16 has 5 blocks. In the block 1 and block 2, they are constructed two convolutional layers and one max pooling layer. In the block 3 to block 5, they are constructed three convolutional layers and one max pooling layer. We used 5 blocks with blocks 1 to block 4 using ImageNet weights and block 5 weights for classifying glioma disease stages. Typically, VGG16 is connected flatten layer, two fully connected layers, and prediction layer after max pooling layer of the block 5. In this project, our CNN model was connected global average pooling (GAP) layer and prediction layer after max pooling of the block 5. We expected that GAP layer reduces the number of weight parameters and enhances overall prediction accuracy.

Using the training data set and validation data set, our model learned for feature selection to classify into each glioma disease stage. The next step, in order to evaluate our constructed model, we inputted test data set into our model. Then the output was evaluated.

### D. Extracting Object-Level-Feature

We approximated several shapes of cell nuclei region to extract cell nuclei region features. Figure 3 shows an example of approximations based on ellipse, convex hull, boundary box, and boundary.

In order to visualize the relationship between the features of cell nuclei shape and the importance region with CNN, we extracted object-level-feature in importance region. Thus, we extracted features following steps.

- Applying Grad-CAM++ to the classification of each disease stage with our CNN model and generating heat map image.
- Extraction Region of insert (ROI) from the heat map images.
- In the process of extracting object-level-features from an important region of CNN decision, we patched(100x100 pixel) the important region so that the region was divided into 9 images like Fig. 4. Also object-level-features were extracted from these patched images.
- We fed the extracted features into a support vector machine (SVM) classifier and reviewed the relationship between the shape of the cell nuclei region and CNN's decision from these results.

### E. Visualization Important Region for Classification

Recently, explainable AI is an important aspect in applying DL models in the medical imaging context, since we can not infer decisions taken by black box models. For instance, Ribeiro et al [10] proposed a local interpretable model agnostic explanations (LIME) approach to confirm importance region by CNN based classifications. Another important work is that of Ramprasaath et al [11] who have proposed gradient camera (Grad-CAM) as explanation method for CNN. Grad-CAM or the technique related to it is often used to understand the interpreter of CNN' s decision in the field of medical image

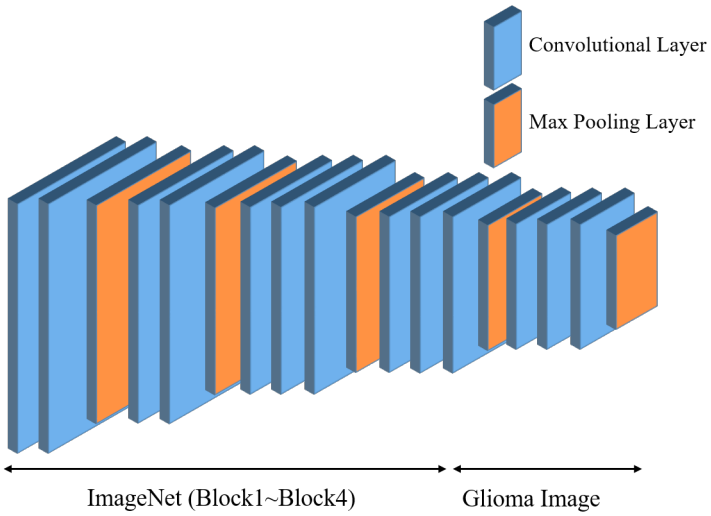
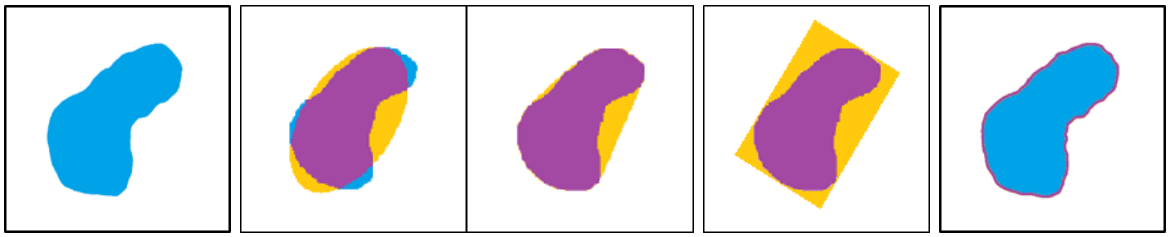



Fig. 2. Our custom designed DL model is based on the VGG16 network. In this work, VGG16 that has 5 blocks was tuned for glioma disease stages. Block 1 to block 4 has ImageNet weight and Block 5 has weight for classifying each glioma disease stage.



(a) Original cell nuclei (b) Ellipse (c) Convex Hull (d) Boundary Box (e) Boundary

Fig. 3. Nuclei object approximations. (a) An example cell nuclei. (b) Ellipse - major axis length, minor axis length, angle, accept ratio, and eccentricity. (c) Convex Hull - area, defect, and solidity (d) Boundary Box - extent, accept ratio. (d) Boundary - perimeter, radii, perimeter curvature.

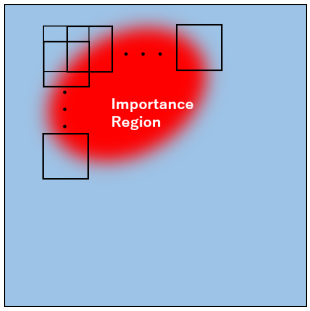
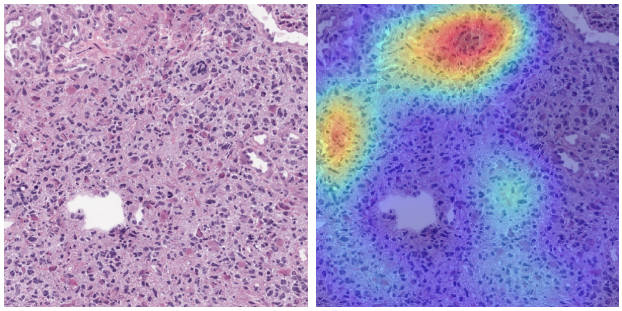



Fig. 4. Method of extracting features from ROI.



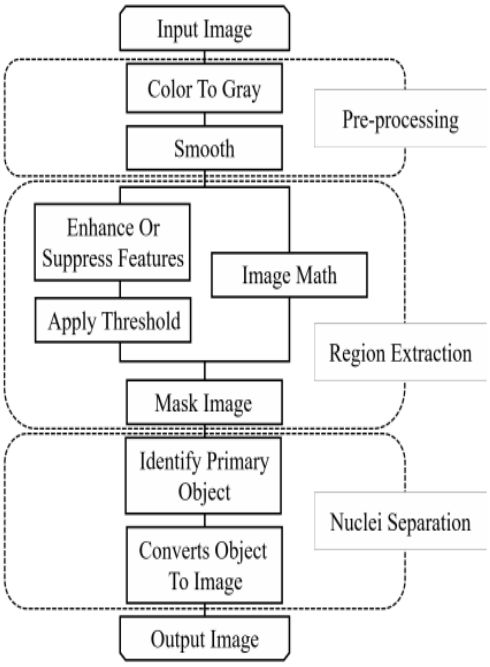
(a) Glioma pathological Image (b) Result of applying Grad-CAM++ for pathological image

Fig. 5. An example of applying Grad-CAM++ technique to our CNN model. (a) Glioma pathological image (GBM), (b) Image visualized importance region by Grad-CAM as heat map. The redder the color of the heat map, the more important that area is in CNN's decision the disease stage of glioma.

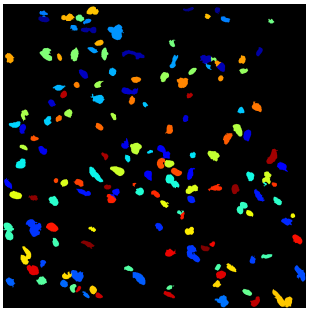
analysis with CNN[12][13][14]. In particular, S. M. Lundberg et al have approached using Grad-CAM to explain CNN's decision in the problem of colonoscopy classification. In this work, we utilized an advanced version named Grad-CAM ++ proposed by A. Chattopadhay [15] to the problem to visualize the region that contributes to CNN decisions like the work of S. M. Lundberg et al. Grad-CAM++ is a popular method that can show us CNN's decision as heat map. And it uses the gradient information flowing into the last convolutional layer of the CNN to understand the importance of each neuron for a decision of interest. Figure 5 shows an example of applying Grad-CAM++ technique to our CNN model.

### *F. Cell Nuclei Segmentation*

We focused on the relationship between CNN's decision and object-level-feature of the cell nuclei. In this paper,




(a) Segmentation pipeline using CellProfiler



(b) Nuclei segmented image

Fig. 6. Cell nuclei segmentation procedure used in our work. (a) Overall segmentation pipeline and (b) example nuclei segmented from an input glioma pathological image.

object-level-feature types of cell nuclei was reference to the features suggested by Gurcan et al.. Thus experimental images need segmentation based on the cell nuclei region. In this paper, these images were run segmentation using CellProfiler (https://cellprofiler.org/) which is an open-source software for measuring and analyzing cell images. The segmentation pipeline of H & E stained glioma cell nuclei has been suggested by Surya et al[5]. Thus in this work, we used the pipeline in literature[5].

## III. Experimental Results and Discussion

For quantitative evaluation of our model, we utilized the following metrics from the classification literature.

TABLE I
CLASSIFICATION ACCURACY FOR GLIOMA PATHOLOGICAL IMAGE.

| | Classification accuracy [%] |
|---|---|
| Dataset 1 | 99.2 |
| Dataset 2 | 99.3 |
| Dataset 3 | 99.2 |
| Average | 99.2 |

TABLE II
CONFUSION MATRIX FOR OUR CNN WITH THE MOST HIGH PERFORMANCE IN THIS EXPERIMENT.

| **Table Head** | **Our Model Prediction** | | |
|---|---|---|---|
| | ***GBM*** | ***LGG*** | ***Recall*** |
| GBM | 995 | 5 | 99.5 |
| LGG | 11 | 989 | 98.9 |
| Precision | 98.9 | 99.7 | |
| F-Measure | 99.3 | 99.3 | |

$$Accuracy = \frac{T_{gbm} + T_{lgg}}{T_{gbm} + F_{gbm} + T_{lgg} + F_{lgg}} \quad (1)$$

$$Precision_{gbm} = \frac{T_{gbm}}{T_{gbm} + F_{gbm}} \quad (2)$$

$$Precision_{lgg} = \frac{T_{lgg}}{T_{lgg} + F_{lgg}} \quad (3)$$

$$Recall_{gbm} = \frac{T_{gbm}}{T_{gbm} + T_{lgg}} \quad (4)$$

$$Recall_{lgg} = \frac{T_{lgg}}{T_{lgg} + F_{gbm}} \quad (5)$$

$$F - Measure = \frac{2Recall * Precision}{Recall + Precision} \quad (6)$$

where $T_{gbm}$ and $T_{lgg}$ they are the number of true predictions by CNN for input test data set. In contrast, $F_{gbm}$ and $F_{lgg}$ are the number of false predictions by CNN for the input test data set. In addition, in order to confirm the accuracy of the learning model, we run an experiment like the above steps a total of three times using the other two data subsets.

### A. Classification accuracy with CNN

The result of learning with deep CNN shows Table I. The classification accuracy was around 99% in every experiment. Table II provides the confusion matrix and evaluation value of CNN model with the highest performance in this paper. In Table II, all evaluation values are around 99%. In particular, the classification accuracy of our model is 99.2% so that it is higher than the classification accuracy of the previous research [8][9]. These results indicate that it was successful to create a learning model for glioma disease stage classification. In following experiments, we used the learning model with the performance showed on Table II.

### B. Relationship between the shape of the cell nuclei region and CNN's decision

We imputed object-level-features into Support Vector Machine (SVM) to confirm the importance of object-level-features by our model. Table III provides the results of

TABLE III
CONFUSION MATRIX USING SVM (KERNEL: 'RBF').

(a) Boundary Feature
($C$ : $5.99 \times 10^{11}$, $\gamma$: $2.78^{-11}$)

| Table Head | SVM Prediction | | |
|---|---|---|---|
| | *GBM* | *LGG* | *Recall* |
| GBM | 124 | 24 | 84 |
| LGG | 65 | 106 | 62 |
| Precision | 82 | 62 | |
| F-Measure | 74 | 70 | |

(b) Bounding Box Feature
($C$ : $6.00 \times 10^{9}$, $\gamma$: $1.29^{-9}$)

| Table Head | SVM Prediction | | |
|---|---|---|---|
| | *GBM* | *LGG* | *Recall* |
| GBM | 118 | 35 | 77 |
| LGG | 647 | 119 | 72 |
| Precision | 72 | 77 | |
| F-Measure | 74 | 74 | |

(c) Convex Hull Feature
($C$ : $1.29 \times 10^{6}$, $\gamma$: $10^{-10}$)

| Table Head | SVM Prediction | | |
|---|---|---|---|
| | *GBM* | *LGG* | *Recall* |
| GBM | 108 | 44 | 71 |
| LGG | 64 | 103 | 62 |
| Precision | 63 | 70 | |
| F-Measure | 67 | 66 | |

(d) Elliptical Feature
($C$ : 6000, $\gamma$:$10^{-8}$)

| Table Head | SVM Prediction | | |
|---|---|---|---|
| | *GBM* | *LGG* | *Recall* |
| GBM | 111 | 35 | 76 |
| LGG | 53 | 120 | 69 |
| Precision | 68 | 77 | |
| F-Measure | 72 | 73 | |

disease stage classification using object-level-features in the importance region by CNN.

In each experiment using outside of the elliptical feature, the trend obtained the best classification accuracy was confirmed when the $C$ was high and $\gamma$ was low comparatively. In the case of the elliptical features, $C$ value was lower than other features. Generally, what the $C$ value was large and $\gamma$ was small means fed data has independent distribution in each class. Thus these results indicate that object-level-features (except for the elliptical features) inside of important regions for CNN's decision have different distributions between LGG and GBM. AS well as above, the elliptical features include many features which do not contribute to classification. Therefore, we consider that there's the relationship between CNN's decision and object-level-features (except for the ellipse features) and the ellipse features need to advanced investigation like feature selection. However, each classification accuracy in Table III is approximately 70%. We consider that these were due to some outlier values. Thus the discussion to reconsider the process of feature extraction and to compare classifiers, to tune parameters will be required. Thus we need to brush up the process of feature extraction and classification to discuss the relationship of CNN's decision more clearly.

Furthermore, CNN decides disease stages from overall features in a given pathological image. In this work, we focused only on object level features. Thus, we need to consider other spatial, or nuclei features in a similar framework. For instance, we are planning to focus on the color feature, positional features of each other cell nuclei.

## IV. CONCLUSION

In this work, we studied how deep CNN's decision to classify each disease stage of brain glioma histopathological images occurs. In particular, we focused on object-level-features of cell nuclei in the importance region visualized by Grad-CAM++. The result of disease stage classification using object-level-features in the importance regions by our CNN model. In each experiment of using Object-Level-Feature, the trend that is obtained the best classification accuracy was confirmed when the C is high and $\gamma$ is low comparatively. These results indicate that object-level-features inside of important regions for CNN's decision basically have different distributions between LGG and GBM. Therefore, we consider that there's the relationship between the CNN's decision and object-level-features. However, each classification accuracy in Table III is around 70%. We consider that these were due to some abnormal values and the lack of feature selection. Thus we need to brush up the process of feature extraction and classification to discuss the relationship of CNN's decision more clearly. In addition, CNN decides disease stages from overall features in given pathological images. Therefore, in the future works, we will focus on the color features and positional features of each other cell nuclei in the importance region visualized by Grad-CAM++. If we are able to classify each grade using its features and visualizing how CNN decides completely, it leads to better decision making systems in the area of computational pathology.

## REFERENCES


[1] Y. LeCun, Y. Bengio, G. Hinton. Deep learning. Nature, 521(7553), pp.436-444, 2015.

[2] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng and M. Chen, "Medical image classification with convolutional neural network," 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), Singapore, 2014, pp. 844-848

[3] A. Kumar, J. Kim, D. Lyndon, M. Fulham and D. Feng, "An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 1, pp. 31-40, Jan.

[4] M. D. Kumar, M. Babaie, S. Zhu, S. Kalra and H. R. Tizhoosh, "A comparative study of CNN, BoVW and LBP for classification of histopathological images," 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, 2017, pp. 1-7

[5] V. B. S. Prasath, K. Fukuma, B. J. Aronow, H. Kawanaka. Cell nuclei segmentation in glioma histopathology images with color decomposition based active contours. IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1734–1736, 2015.

[6] K. Fukuma, V. B. S. Prasath, H. Kawanaka, B. J. Aronow, and H. Takase. A study on nuclei segmentation, feature extraction and disease stage classification for human brain histopathological images. 20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES), pp. 1202–1210, 2016.

[7] A. Yonekura, H. Kawanaka, V. B. S. Prasath, B. J. Aronow, H. Takase. Glioblastoma multiforme tissue histopathology images based disease stage classification with deep CNN. International Conference on Informatics, Electronics & Vision (ICIEV), Hyogo, Japan. Proc. IEEE, Sep 2017.

[8] A. Yonekura, H. Kawanaka, V. B. S. Prasath, B. J. Aronow, H. Takase. Automatic disease stage classification of glioblastoma multiforme histopathological images using deep convolutional neural network. Biomedical Engineering Letters, 8(3), pp. 321-327, 2018.

[9] M. G. Ertosun, D. L. Rubin, 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, 2015, 1899–1908. Published online 2015 Nov 5.

[10] M. T. Ribeiro, S. Singh, C. Guestrin. Why should I trust you? Explaining the predictions of any classifier, eprint arXiv:1602.04938, 2016.

[11] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. InProceedings of the IEEE international conference on computer vision 2017 (pp. 618-626).

[12] S. Hicks et al., "Dissecting Deep Neural Networks for Better Medical Image Classification and Classification Understanding," 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), Karlstad, 2018, pp. 363-368.

[13] G. Zhao, B. Zhou, K. Wang, R. Jiang, M. Xu, Respond-CAM: Analyzing Deep Models for 3D Imaging Data by Visualizations. In: Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, vol 11070. Springer, Cham, 2018.

[14] S. Lee, J. Lee, J. Lee, C.-K. Park, S. Yoon. Robust tumor localization with pyramid grad-cam. arXiv preprint arXiv:1805.11393, 2018.

[15] A. Chattopadhay A. Sarkar, P. Howlader, V. N. Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In IEEE Winter Conference on Applications of Computer Vision (WACV) 2018 Mar 12 (pp. 839-847). IEEE.

[16] M. N. Gurcan, L.E. Boucheron, A. Can, A. Madabhushi N. M. Rajpoot, B. Yener. Histopathological image analysis: A review, IEEE Reviews In Biomedical Engineering, vol. 2, pp. 47–171, 2009.