

Deep Learning Use for Differentiation of Low-grade vs High-Grade Glioma in Intraoperative Squash Smears

Bhushan Diwakar Thombre

NIMHANS: National Institute of Mental Health and Neuro Sciences

B N Nandeesh

NIMHANS: National Institute of Mental Health and Neuro Sciences

Vikas Vazhayil

NIMHANS: National Institute of Mental Health and Neuro Sciences

A R Prabhuraj (✉ drprabhuraj@yahoo.co.in)

NIMHANS: National Institute of Mental Health and Neuro Sciences <https://orcid.org/0000-0002-5986-1521>

Research Article

Keywords: deep learning, artificial intelligence, glioma, squash smear, convolutional neural network, feature visualization

Posted Date: March 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-329319/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objective

Automated diagnosis using Artificial Intelligence (AI) techniques would be a useful addition to the intraoperative squash smear diagnosis. A robust diagnostic tool would enhance capabilities in centres where there is limited expertise for the diagnosis of intracranial lesions. The study aims to explore possibilities of deep learning technique-based models to classify squash smear images of glioma into high- and low-grade tumors.

Methods

500 Scanned images of squash smear were obtained intraoperatively and dataset was built. Image dataset was then pre-processed and fed into a CNN (Convolutional Neural Network) model for training and validation. The dataset consisted of 10,000 images of high (6000) and low (4000) grade gliomas, divided into three sets of training, validation and testing.

Results

CNN model based on deep learning algorithm was built and trained on training dataset to get accuracy of 96.2%. On a testing dataset which contains images previously unseen by trained model, it could achieve accuracies of 91% for diagnosing high grade glioma and 77% for low grade glioma. A positive predictive value of 86.6% and F1-score of 0.887 was achieved. Feature visualization technique was applied at the end to visualize regions of interest.

Conclusion

Deep Learning techniques can be applied as diagnostic tool if proper standardized images are obtained for reporting of squash smears of gliomas. The diagnostic accuracies of such tools can reach up to current standard diagnostic accuracies by conventional ways of reporting. Feature visualization techniques applied which can be used for rapid screening of slides or section of slide to assist in rapid diagnosis.

Introduction

Central Nervous System (CNS) malignancy accounts for nearly 1.3 - 2% of all tumors in India and worldwide has an incidence of 5-10 per 1,00,000 persons[1-3]. Computer Aided Diagnosis (CAD) has been used previously in many fields related to neurosurgery. In neuroradiology, for analysis of images acquired via various modalities like X-ray, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), etc. have been used for rapid and better diagnosing processes. Artificial Intelligence has been used in report generation based on CT in Traumatic Brain Injury⁴, in segmentation of morphology of brain tumor and differentiation of high-grade and low-grade brain tumor based on MRI[5]. Many of these approaches were

heavily dependent on feature extraction from the images and appropriate image pre-processing, which can be a tedious task and are prone to errors.

Intraoperative squash smear cytological preparation was first introduced by Eisenhardt and Cushing in the early 1930 and Badt in 1937.[6] Diagnostic accuracies of squash smears range from 76 – 96 % in different studies.[6] Typically, this procedure takes around 30 minutes to 2 hours at various centers. At our center the typical turnaround time for diagnosis is in-between 30 to 45 mins. This situation gets added difficulties during emergencies and in wee hours, when an expert neuropathologist may not be available. At some places, the facility of intraoperative diagnostics may not be available at all due to a lack of expertise. Also, there is no readily available tool to confirm quality of tissue obtained like during framed/frameless stereotactic biopsy.

Artificial Neural Network (ANN) based techniques can be used to address at least some of the issues mentioned earlier. Deep learning, a form of complex ANN, is getting recently very popular in solving many challenges in trend analysis, computer vision analysis, computer-aided diagnosis etc. Also, this method can provide some certainty for neurosurgeons that the tissue sent for final analysis is indeed a representative tissue of tumor during blind procedures like stereotactic biopsies. It can also facilitate repeatability of samples on the operating table itself. To our best knowledge, the use of CNN on brain squash images has never been tried before.

Methods

The aim of this study was to build a CNN model specific for the analysis of the squash smear cytology of the brain tumor tissue and to check the validity of such AI Agents in detecting high-grade vs low-grade gliomas.

After preparing the dataset from smear cytology slides, a deep learning CNN Model designed specifically for handling such medical images was developed. Various steps in the process are described below.

Image Acquisition and Data Collection

Approval from Institutional Ethics Committee was obtained before collecting the data. Intraoperatively prepared squash smears slide of gliomas were acquired for surgeries done for 18 months period from June 2017 to December 2018 of gliomas from Department of Neuropathology. Of 500 slides scanned, an average of 20 representative images were obtained from each of the smear slides. Only non-overlapping uniformly spread, single cell layer parts and well stained areas in the slide were selected for imaging by expert neuropathologist. These images have to be selected this way for training purposes as in neural network training input data has to be labelled correctly. The ground truth was acquired from the final histopathology report.

These slide images were then converted into digital format by digital microscope (Olympus® Bz53, DP27 Camera). The data collected for the study were images of the squash smear slides whose final

histopathology diagnosis was glioma. Images obtained were then divided into three sets, a training set, validation set and testing dataset (Figure 1). Images in the training set were only used to train the network and the rest of the images were only used for validation/testing. Of the total 10,000 images collected, 6,000 belonged to High Grade Glioma and 4,000 belonged to Low Grade Glioma. Amongst them, 3,200 were used for training and rest for testing/validation. Only high- and low-grade glioma were taken into consideration for this initial project as on squash smear or frozen section many times even by expert neuropathologist this much granularity can be obtained for gliomas and immunohistochemistry has to be done later during final histopathology reporting.

Image Preprocessing

The images acquired were of height of 480 pixels and a width of 612 pixels, coded in RGB Standard Code, and coded in 3 channels (612 x 480 x 3). The first image resolution was reduced to (150 x 150 x 3) for computing. After that image was vectorized into linear arrays on each channel. The arrays are then normalized with Min-Max Scaler Normalization and used as inputs. The input of the images was then processed through the Image Data Generator, to add randomness, noise, rotations, and other parameters for making data more generalized. Figure 2 demonstrates image before pre-processing (a) and after (b) pre-processing.

Building of CNN Model

Convolutional Neural Network (CNN) are standard deep learning methods to work with image data. CNN uses a convolution function mentioned below was employed:

$$Y_k = f(W_k * x)$$

This formula is an oversimplification of the actual convolution formula⁷, but in scope of this article, we consider it to fulfill its purpose. Here, the inputs are denoted by x , the filter for the k^{th} feature map is denoted by W_k . The $f(\cdot)$ function denotes the Convolutional function and the Y_k denotes the output of the function given the input x at the k^{th} position. Convolution itself is a linear operator at its core.

A CNN model was built based on VGG19⁷ with added layers at the end which were specially curated to handle these processed images (Figure 3). Models were built in Python language with Keras library with TensorFlow as backend which all are open source packages. As can be seen in the below mentioned figure, the input is in the shape of (150x150x3) formation. The layers were later built as alternate layers of Convolutional function, as mentioned previously, with Pooling layers which are connected with each other via Dense connecting layers. Once the architecture of the model is built; on the top of the model another set of ANN is built and connected and the final 5 layers of the network are formed. These layers were specifically built considering the input and nature of data. Model is finally compiled with "Stochastic Gradient Descent (SDG)" with loss of function as "binary cross-entropy" and metrics of accuracy as "accuracy" and a total of 28,939,329 trainable parameters.

The final output layer was considered as only binary output and labelled as either '1' for high-grade glioma or '0' for low-grade gliomas. This layer has the activation function of "sigmoid", which gives output probabilities between 0 and 1. This would provide the confidence in the probability of high-grade glioma or low-grade tumors in the CNN network's prediction.

Training of Model

Afterward, the CNN Model was trained on a workstation with 16 GB RAM, the process augmentation was done with NVIDIA® RTX 2060 6 GB Graphic Processor, on Intel® Architecture which took 121.02 minutes to train in a batch of 32 images with 100 epochs. One should note that all of the training is done without any feature extraction and no human intervention whatsoever, also for actual reporting purposes much lower configuration of the computer system should suffice.

Validation and Statistics

The artificial neural network was trained using these images on the training set and the accuracies and cross-validation matrices were built. This would help in validating the fitness of the model for the generalized use and for practice in the real world. Statistical analysis was done by Scikit-Learn Library, which is integrated with Keras library.

Results

A) Results During Training the Dataset:

As shown in Figure (4), which is a line chart for Training and Validation Data loss and accuracies, for loss the lesser value is considered better and for accuracies higher is regarded better. It was noted that the model started converging at around 20th epoch and with minimal overfitting. Specific techniques like Learning rate regulation and Dropouts were used for to avoid overfitting. Also, multiple models with various hyperparameters were tested before finalizing the best working model. The task of hyperparameter tuning for the purpose was tedious and time consuming. On reaching 100th epoch, the loss was 0.0950 in the training set and 0.1016 for the validation set. Accuracies of 96.2% were reached on the training set and 96.39% on the validation set. The validation data set was used only for internal comparison and not for training the model.

B) Results During Testing phase:

The testing dataset was used for obtaining results on final data which was not previously exposed to the model. Results were obtained in form of '1' for "high-grade glioma" and '0' for "low-grade glioma" as mentioned before.

Confusion Matrix is a cross table between true label versus predicted label which is 2x2 table as shown in Figure (5). Sensitivity and Specificity as well as Positive and Negative Predictive values were calculated as demonstrated in Table 1. As indicated in the table, the accuracy in prediction for high grade glioma

was 91% and for low grade glioma it was 77%. These reports could be generated in fraction of time on each image.

The F1-Score which is a harmonic means of precision and recall gave the overall performance of the model. Highest value of the score is considered 1 and lowest being 0. The F1-Score for our model was 0.887.

C) Feature Visualization:

Figure (6) shows how the network sees each slide image. But this is of little use to humans. A method called 'feature visualization' can be used to make images human compatible. A heatmap of combined features were overlaid on the original image to be analyzed and shown in Figure (7). This could help in screening thousands of high-power fields in a given slide in limited time to help automate screening of slides.

Discussion

Squash smear is prepared from tumor tissue obtained during surgery and sent for analysis to get the idea about the nature and type of the tumor a neurosurgeon is dealing with. The accuracy in various studies comparing squash diagnosis with that of the final histopathological diagnosis ranges from 83% to 95%. There are variable accuracies in reporting squash obtained from stereotactic biopsy procedures and direct tumor decompression procedures.[6,8-10]

Artificial Intelligence (AI) can be used to assist in various diagnostic methods. Previous studies which were done with some form of feature selection application indicated that the cervical cancers could be diagnosed with accuracies ranging from 85% to 90% on external testing dataset.¹¹ Also, in cases of Thyroid Cancer, fine needle aspiration cytology diagnosis could be done with sensitivity of 90.48%.[12] But in many of these studies, the sample size in all these studies very limited, and the generalization would be doubtful.

In neuroradiology, Sasank Chilamkurthy, et al[14] demonstrated the utility of CNN networks to predict different abnormalities in various scenarios from non-contrast CT scans. This study has shown that accuracies up to 92% can be achieved in detecting pathologies like intracranial hematoma or subdural or extradural hematoma. Their network could also predict bony abnormalities like skull fractures (92.0%), midline shift and mass effect (93.0%).

In CNS histology, classification of histopathological slides for CNS tumor and segmentation was done by a similar use of CNN. In a paper by Yan Xu, et al[13] they have used 23 Histopathological slides of Glioblastoma Multiforme and 22 images of low-grade glioma and it could achieve accuracies of up to 97.5%. However, in our study, we used 10,000 squash smear images for building and validating the CNN model. A specific study diagnosing malignant vs non-malignant breast cancer based on Computer-aided histological diagnosis could achieve an accuracy of 88.3%.[14]

In the present study, sensitivity for detection of high-grade glioma and low-grade glioma by CNN AI agent was 91% and 77% which were comparable to conventional detection methods by human pathologists (83-95%)^{6,11} and 77-80%^[15] in various studies.

Feature visualization can help us visualize the inner workings of these generally considered black box models. It gives us an idea about which parts of the image our model is giving importance to, while making the diagnosis. If this network is used to screen the whole slide it can give the interpreter an overview of relevant regions of interest. Screening whole slides with feature visualizing techniques can reduce reporting time where experts are available.

Other similar techniques are likely on rise like diagnosing with Raman spectroscopy which has decades of research behind it^[16,17], but as compared to our method it requires specialized equipment for reporting.

All the computer programming work done in this study was by team members only. Promoting learning of computer programming languages in the field of medicine is a need of the hour in recent future.

Future Possibilities and Limitations:

Future holds many possibilities for intraoperative tissue diagnosis reporting. The same model can be made more accurate and more generalized for the diagnosis of all types of CNS pathologies with larger datasets. Similar models can be useful in the diagnosis of the final histopathological slides once trained. Also, if implemented with surgical confocal microscopes, it can help delineate the normal from abnormal brain tissue while operating in real time. We will try to achieve the same and build better models in the future and make it available publicly to assist neurosurgeons and neuropathologists.

Although the reporting accuracies for low grade gliomas are comparable to human standards, there is scope of improvement with better upcoming deep learning techniques. In this study we tried to demonstrate the proof-of-concept implementation of diagnostic intraoperative neuropathology.

Conclusion

We have demonstrated that deep learning models can be used for diagnosis of high- and low-grade gliomas in squash smear pathology.

Artificial Intelligence can be reliably used, if properly standardized images of CNS Tumor squash smear cytology are obtained. In our study we found that a CNN Model can differentiate and diagnose High Grade Gliomas from Low Grade gliomas efficiently with accuracies of 91% and 77% respectively which is comparable to current human diagnostic accuracies. Also, feature visualization tools based on these models can screen large areas on slides for further detailed analysis by an expert neuropathologist. We strongly believe with even larger datasets the generalizability of such models can improve in future.

These results will promote to do future research in intra-operative neuropathology to differentiate other types of clinically relevant tumors like germinoma, lymphomas, types of meningiomas and more.

Declarations

Disclosures:

No conflict of interests to disclose.

Acknowledgments: Dr. Pooja Hazare for proofreading and diagrams within the paper.

Ethics Approval: Taken from by NIMHANS ETHICS COMMITTEE (No.NIMH/DOIED(BS & NS DIV)/2017-18

References

1. Nair M, Varghese C, Swaminathan R. Cancer: Current scenario, intervention strategies and projections for 2015. *Burd Dis India*. Published online 2015.
2. Yeole BB. Trends in the brain cancer incidence in india. *Asian Pac J Cancer Prev*. 2008;9(2):267-270.
3. Brain and Other Nervous System Cancer - Cancer Stat Facts. Accessed January 10, 2018. <https://seer.cancer.gov/statfacts/html/brain.html>
4. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *The Lancet*. 2018;392(10162):2388-2396. doi:10.1016/S0140-6736(18)31645-3
5. Emblem KE, Nedregard B, Hald JK, Nome T, Due-Tonnessen P, Bjornerud A. Automatic glioma characterization from dynamic susceptibility contrast imaging: brain tumor segmentation using knowledge-based fuzzy clustering. *J Magn Reson Imaging JMRI*. 2009;30(1):1-10. doi:10.1002/jmri.21815
6. Jindal A, Diwan H, Kaur K, Sinha VD. Intraoperative Squash Smear in Central Nervous System Tumors and Its Correlation with Histopathology: 1 Year Study at a Tertiary Care Centre. *J Neurosci Rural Pract*. 2017;8(2):221-224. doi:10.4103/0976-3147.203811
7. Rawat W, Wang Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput*. 2017;29(9):2352-2449. doi:10.1162/neco_a_00990
8. Cappabianca P, Spaziante R, Caputi F, et al. Accuracy of the analysis of multiple small fragments of glial tumors obtained by stereotactic biopsy. *Acta Cytol*. 1991;35(5):505-511.
9. Mitra S, Kumar M, Sharma V, Mukhopadhyay D. Squash preparation: A reliable diagnostic tool in the intraoperative diagnosis of central nervous system tumors. *J Cytol Indian Acad Cytol*. 2010;27(3):81-85. doi:10.4103/0970-9371.71870
10. Bora K, Chowdhury M, Mahanta LB, Kundu MK, Das AK. Pap smear image classification using convolutional neural network. In: *Proceedings of the Tenth Indian Conference on Computer Vision,*

Graphics and Image Processing - ICVGIP '16. ACM Press; 2016:1-8. doi:10.1145/3009977.3010068

11. Miranda GHB, Barrera J, Soares EG, Felipe JC. Method for Computational Analysis of Histopathological Images to Support the Diagnosis of Cervical Cancer. :6.
12. Sanyal P, Mukherjee T, Barui S, Das A, Gangopadhyay P. Artificial Intelligence in Cytopathology: A Neural Network to Identify Papillary Carcinoma on Thyroid Fine-Needle Aspiration Cytology Smears. J Pathol Inform. 2018;9. doi:10.4103/jpi.jpi_43_18
13. Xu Y, Jia Z, Ai Y, Zhang F, Lai M, Chang EI-C. Deep convolutional activation features for large scale Brain Tumor histopathology image classification and segmentation. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2015:947-951. doi:10.1109/ICASSP.2015.7178109
14. Arau T, Aguiar P, Eloy C. Classification of breast cancer histology images using Convolutional Neural Networks. Published online 2017:1-14.
15. Roessler K, Dietrich W, Kitz K. High diagnostic accuracy of cytologic smears of central nervous system tumors. A 15-year experience based on 4,172 patients. Acta Cytol. 2002;46(4):667-674. doi:10.1159/000326973
16. DePaoli D, Lemoine É, Ember K, et al. Rise of Raman spectroscopy in neurosurgery: a review. J Biomed Opt. 2020;25(5):050901. doi:10.1117/1.JBO.25.5.050901
17. Hollon TC, Pandian B, Adapa AR, et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. Nat Med. 2020;26(1):52-58. doi:10.1038/s41591-019-0715-9

Tables

Table 1: Table showing sensitivity, specificity, positive and negative predictive values for the patient. Also note the F1-score is mentioned

PERFORMANCE OF MODEL ON TESTING DATASET	
True Positive Rate (Sensitivity)	91%
True Negative Rate (Specificity)	77%
Positive Predictive Value (PPV)	86.6%
Negative Predictive Value (NPV)	83.89%
F1-Score	0.887

Figures

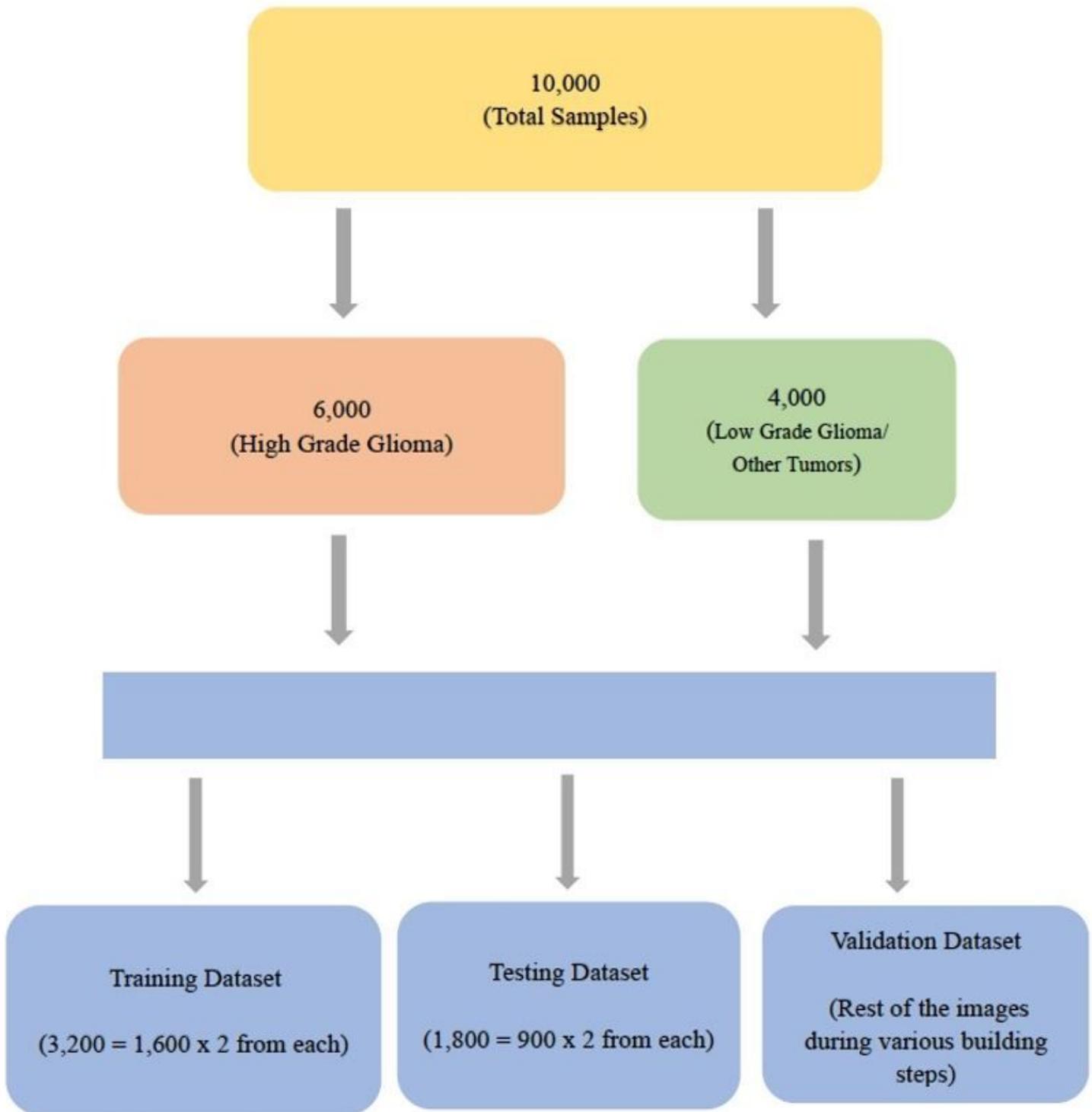
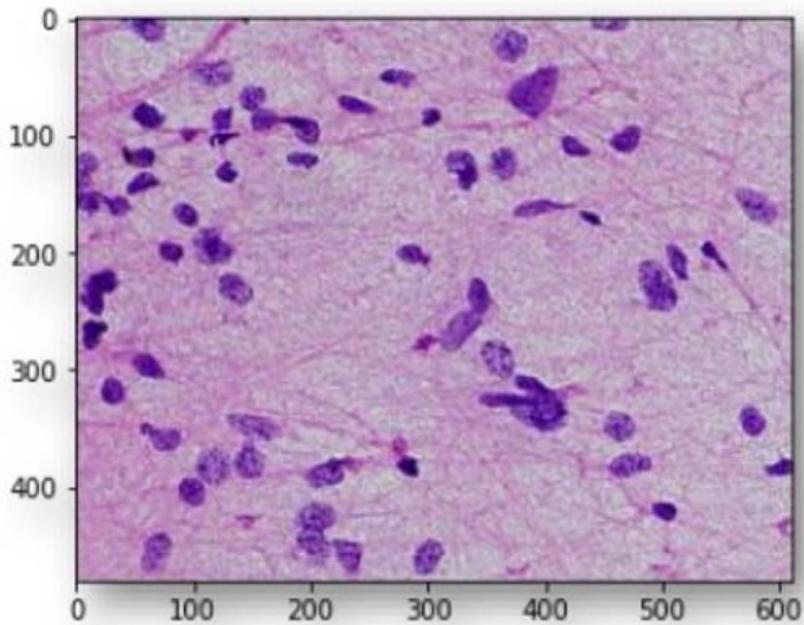
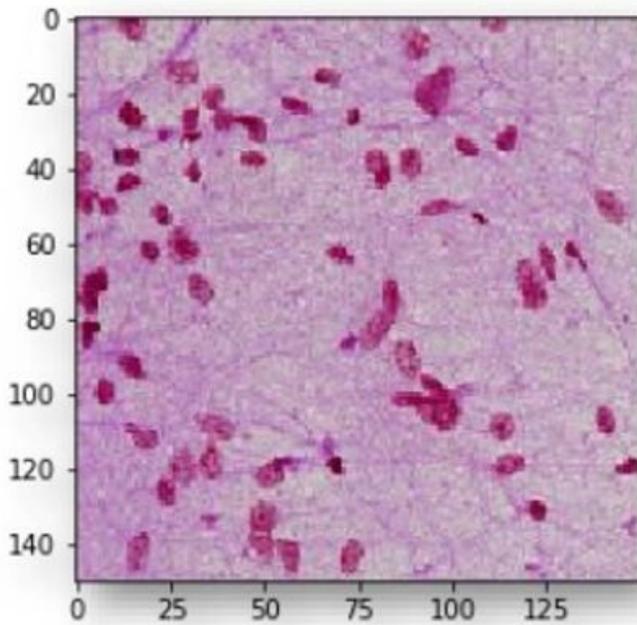


Figure 1

Distribution of images in different dataset done randomly. Note that the rest of the data is used for the testing and validation for the model, this is the data which is never seen by the model during training period and expected to help in generalization of the model



(a) Image before Pre-processing



(b) Image after Pre-processing

Figure 2

Image before(a) and after(b) pre-processing is done. Some colour changes also might be noted due to pre-processing algorithm

Layer (type)	Output Shape	Param #
input_1 (Input Layer)	(None, 150, 150, 3)	0
block1_conv1 (Conv2D)	(None, 150, 150, 64)	1792
block1_conv2 (Conv2D)	(None, 150, 150, 64)	36928
block1_pool (MaxPooling2D)	(None, 75, 75, 64)	0
block2_conv1 (Conv2D)	(None, 75, 75, 128)	73856
block2_conv2 (Conv2D)	(None, 75, 75, 128)	147584
block2_pool (MaxPooling2D)	(None, 37, 37, 128)	0
block3_conv1 (Conv2D)	(None, 37, 37, 256)	295168
block3_conv2 (Conv2D)	(None, 37, 37, 256)	590080
block3_conv3 (Conv2D)	(None, 37, 37, 256)	590080
block3_conv4 (Conv2D)	(None, 37, 37, 256)	590080
block3_pool (MaxPooling2D)	(None, 18, 18, 256)	0
block4_conv1 (Conv2D)	(None, 18, 18, 512)	1180160
block4_conv2 (Conv2D)	(None, 18, 18, 512)	2359808
block4_conv3 (Conv2D)	(None, 18, 18, 512)	2359808
block4_conv4 (Conv2D)	(None, 18, 18, 512)	2359808
block4_pool (MaxPooling2D)	(None, 9, 9, 512)	0
block5_conv1 (Conv2D)	(None, 9, 9, 512)	2359808
block5_conv2 (Conv2D)	(None, 9, 9, 512)	2359808
block5_conv3 (Conv2D)	(None, 9, 9, 512)	2359808
block5_conv4 (Conv2D)	(None, 9, 9, 512)	2359808
block5_pool (MaxPooling2D)	(None, 4, 4, 512)	0
flatten_1 (Flatten)	(None, 8192)	0
fc-1 (Dense)	(None, 1024)	8389632
dropout_1 (Dropout)	(None, 1024)	0
fc-2 (Dense)	(None, 512)	524800
dropout_2 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 1)	513

Total params: 28,939,329
Trainable params: 28,939,329
Non-trainable params: 0

Figure 3

CNN Model based on VGG19 with additional layers. All parameters were re-trained to curate results for the squash images

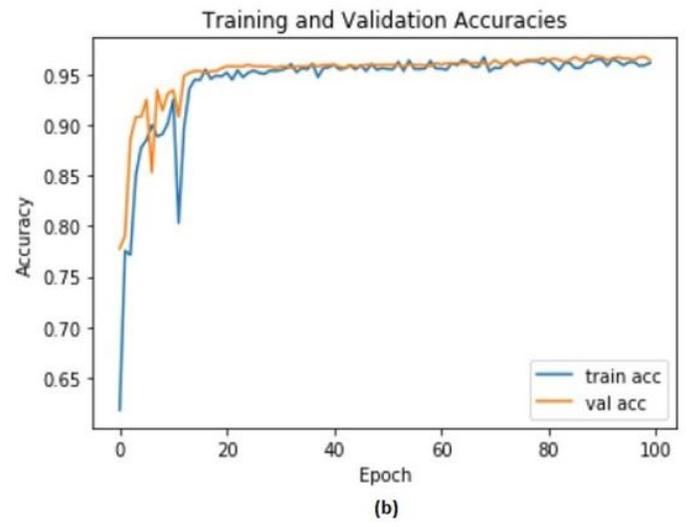
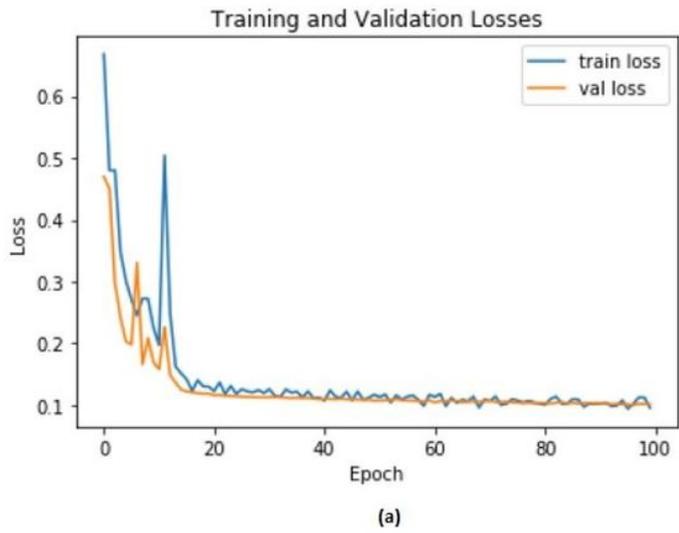


Figure 4

Training and Validation (a) Loss and (b) Accuracies over 100 training epochs

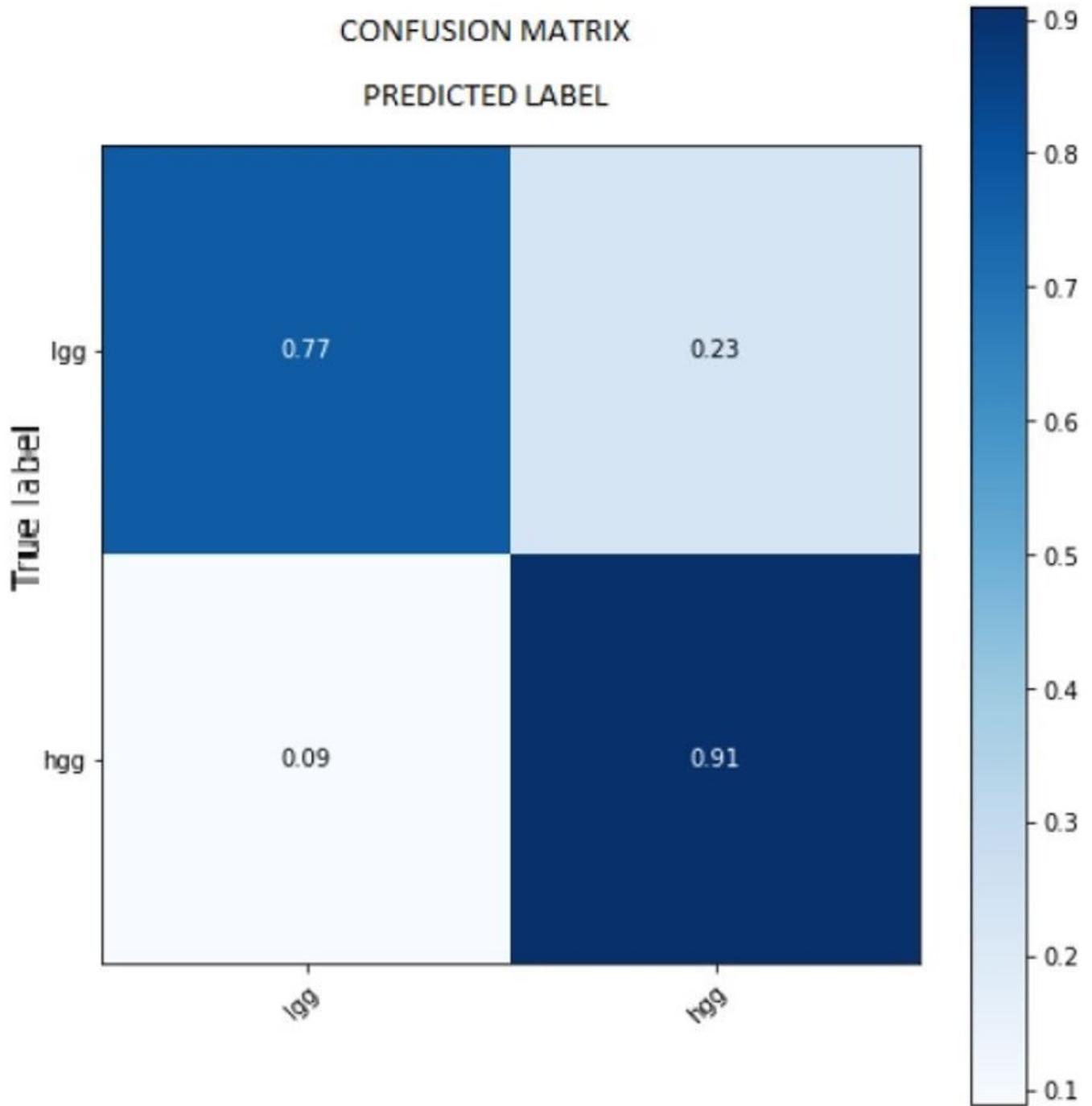
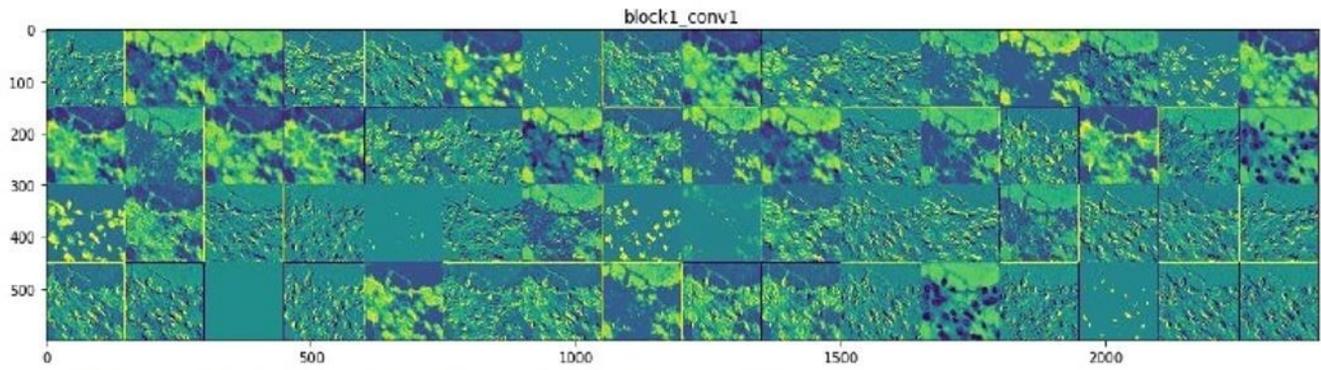
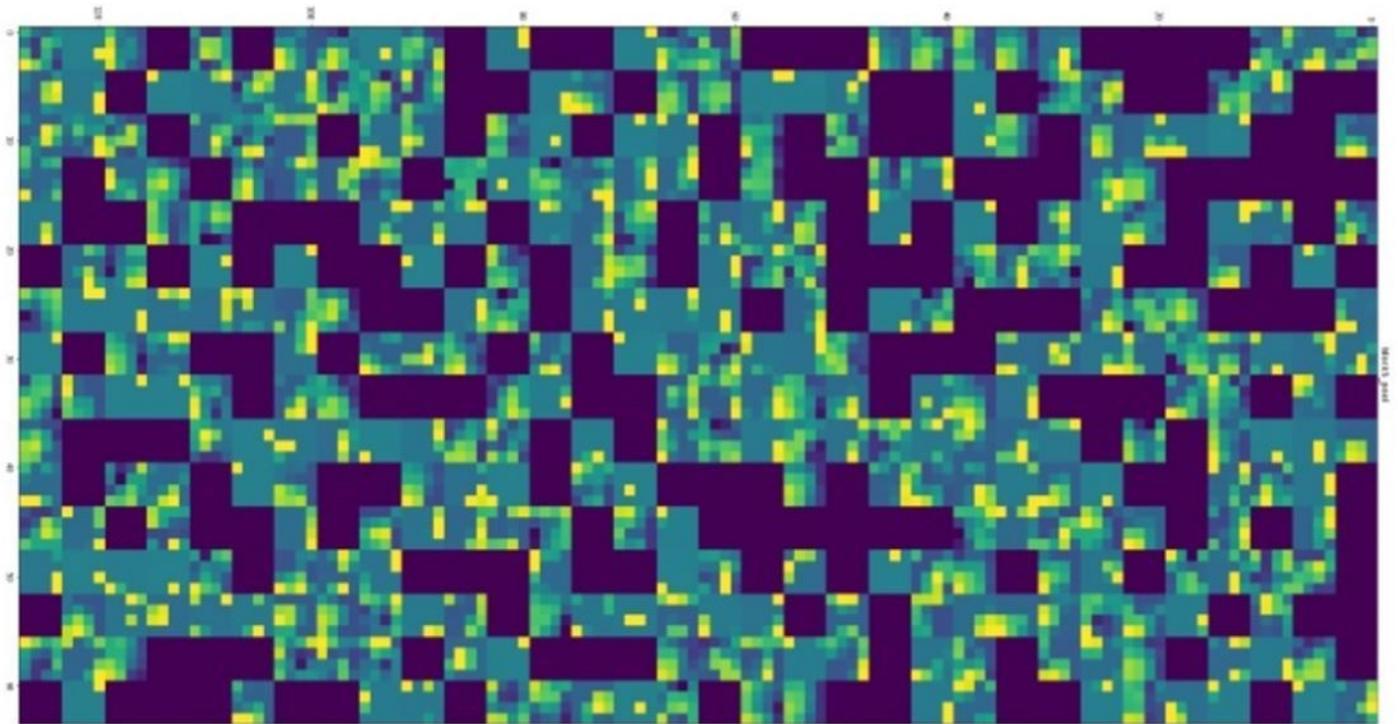


Figure 5

Confusion Matrix built on the testing dataset. (NB:hgg = High Grade Glioma, lgg = Low Grade Glioma)



(a) First Layer Visualized(*block1_conv1*)



(b) Last Layer Visualized(*block1_conv5*)

Figure 6

Various stages of images seen by model. (a) Original Image given as input to the CNN Model (b) First Layer (*block1_conv1*) Parameter Visualization

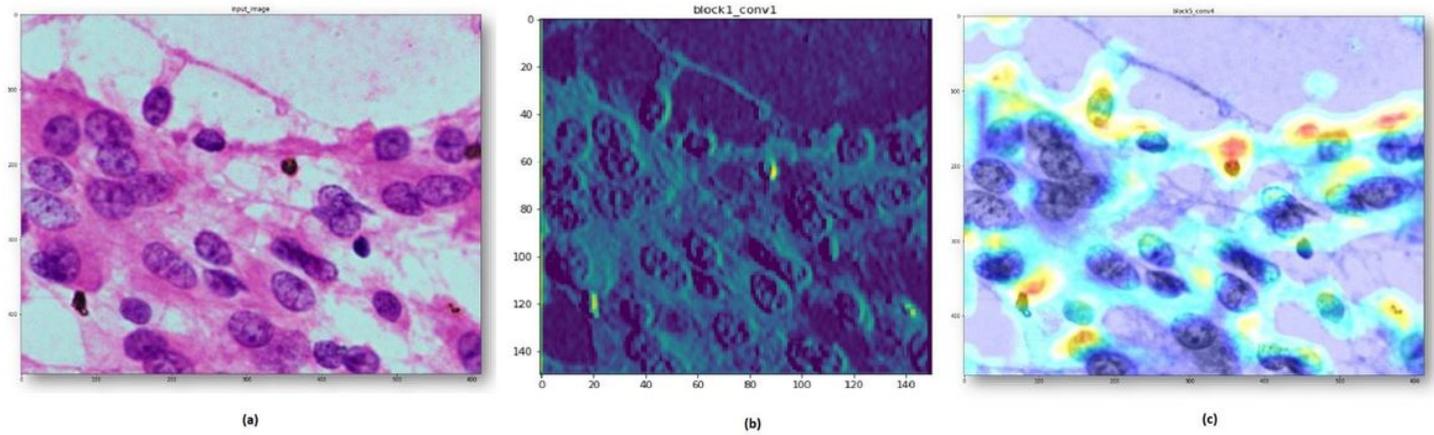


Figure 7

First(a) and Last(b) layers of network. The image demonstrates how the image is actually seen by the network in the first and last layer of the network (c) Heatmap of the Output from the final layer(block5_conv5) obtained from CNN Model is superimposed on the greyscale image of the original image in background. The red colour indicates where the maximum importance is given, while making the decision that it is High-Grade Glioma, followed by yellow and green. The background and normal cells are given the least importance in making the final decision by our model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [annotation.docx](#)