

Glioblastoma Multiforme Classification by Deep Learning Techniques on Histopathology Images

P. Sobana Sumi, Radhakrishnan Delhibabu



Abstract: Brain tumor is one of the most dangerous diseases, which is very hard to diagnose due to its rare symptoms. Diagnosing disease at right time helps to give proper treatment and could extend patient survival period. Histopathology images of brain tumor are taken from the Cancer Genome Atlas (TCGA). Large numbers of tissues have to be analyzed to diagnose disease efficiently, which produces time consuming problem. In this model CNN architecture like InceptionV3 and InceptionResNetV2 are adapted to solve binary and multi-class issues in brain tumor histology images using transfer learning. InceptionV3 is used to extract features and for fine-tuning, InceptionResNetV2 is used for feature extraction. Framed autoencoder network to transform the extracted features to low dimension space and to do clustering analysis on image. Proposed autoencoder produce better clustering result than features extracted by InceptionResNetV2.

Keywords: Glioma, Histology, Transfer Learning, InceptionResNetV2, Classification.

I. INTRODUCTION

Brain tumor turns into cancerous tumor (malignant) or to be found normal tumor (benign) based on its characteristics. Glial (malignant) tumor has these following stages Astrocytomas, Ependymomas, GBM (Glioblastoma Multiforme), Medulloblastomas and Oligodendroglioma. Histopathology image analysis is the way to diagnose surgical tissue (biopsy) by radiologist. Under various microscopes the tissues are diagnosed. This standard method of diagnosing finally produces detailed information about cancer disease or some other disease (present stage). For diagnosing a single case, analysis has to be done on several tissues with multiple staining. TCGA (The Cancer Genome Atlas) contains tissue images of brain tumor which includes Lower Grade Glioma (LGG) and High Grade Glioma (HGG). LGG includes grades 1 and 2 (Astrocytomas, Ependymomas) and HGG includes grades 3 and 4 (Glioblastoma Multiforme, Medulloblastomas, Oligodendrogliomas). In these types of tumors, nuclei images were stained deep purple and other tissues were stained pale purple and red according to H and E staining method. GBM is often referred to as grade IV Astrocytoma which is very hard to detect and its disease stage.

Astrocytoma develops from star shape glial cell called astrocytes, which act as brain support tissue. They are found frequently in brain, mainly in cerebrum. Adults, middle-aged men get affected by this kind of tumor, but rarely found in young people and children. Ependymomas are formed in ependymal cells which are found in the line of ventricular system. Due to its neoplastic transformation, ependymomas is formed. GBM is high grade glioma, this grow faster and spread on other tissues. GBM is commonly found in men and women around the age of 50 to 70. Medulloblastoma grow on cerebellum, found in children. Chemotherapy and radiation helps to cure this. Oligodendrogliomas are formed in myelin. Myelin work as insulation for neuron interconnections in brain. Histopathology tissue analysis is completely a manual way of analysis and it leads some problem due to large amount of patient's data. This cause's huge burden for pathologist in analyzing also this produces time consuming problem. Because of this drawback, new method has to be developed that helps to overcome all these process. For detecting tumors in histopathology tissue images, computer-assisted diagnosis (CAD) systems were used. In CAD technique, first the images are scanned, and then the digital images are processed and analyzed by computer based methods like feature extraction and with machine learning techniques. Due to various types of scanners, staining procedure, age variations and by tissue thickness, digital pathology images have strong color differences. Color normalization is required because it is very hard to classify colors among samples. Feature extraction and classification process is needed for diagnosing patients [1]. For detecting specific diseases like breast cancer (or) esophagitis CAD technique can be used. But this CAD technique is not good for brain tumor analysis. Glioma is a malignant type tumor. Efficient feature descriptor is used for glioma histopathology images [2]. Automatic image analysis technique is developed for nuclei segmentation, feature extraction and disease stage classification. K-S test is used to assure automatic nucleus segmentation, feature extraction and tested images are classified by Support Vector Machine (SVM). This kind of feature descriptor is not sufficient to assure good result [3]. Here used K-S test results on Object-Level features and Spatial-Arrangement features. SVM in case of object level features are higher than SVM in case of spatial arrangement features. Combination of feature descriptor will provide accuracy in classification but this leads to tedious process [4]. Deep learning is a technique that performs classification and feature description simultaneously for glioma histopathology images. In typical classification and pattern recognition, CNN takes an image as input and produces a probability map as output. It performs multiple operations through hidden layers.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

P.Sobana Sumi^{*}, School Of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India.

Radhakrishnan Delhibabu^{*}, School Of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

CNN operations are convolution, pooling, rectified linear unit (ReLU) and softmax. Convolution layer convolve image with filter that consist of certain weights to produce activation map. Low level features such as curves, lines, edges are detected first and later at the end high level features are generated. Pooling perform down sampling and sub-region summarization. ReLU is a non-linear function, which solve gradient vanishing problem. Also, its sparse representation which makes network to learn fast. Softmax take vectors and passes it to fully connected layer that produces probabilistic values based on classification result. CNN architecture, with MxNet framework which consists of 2 convolution layers, 2 pooling layers and 2 tanh. This network is not deep enough to capture the properties of our complex Glioma histopathology images [5]. Developed deep CNN architecture which consists of 3 convolution layers, 3 pooling layers and 3 tanh. The filter size is 5*5 pixels for the convolution layers, and then 2*2 pixels for the pooling layers, last two are fully connected layer. The last softmax layer is used to determine the difference between two classes and to produce a patch that defines whether it is a GBM or LGG. Classification accuracy of the proposed CNN is lower than the previous CNN. It is proposed only for disease stage classification of brain tumor histopathology images. Therefore, it is not necessary to compare itself with the classification accuracy [6]. This method describe deep convolution neural network (Deep CNN). It consists of 7 convolution layers, 8 ReLU and 6 pooling layers. The filter size is 5*5 for convolution layer, 3*3 for pooling layer and end with softmax layer. This gives high accuracy in classification [7]. Though there are issues in identifying features and disease stage classification, another challenge is training labelled data which are less in number. Less number of label data is a common problem in medical field analysis. Using Transfer learning, feature identification and lack of training data problems are solved. Our proposed system uses CNN architecture of InceptionV3 and InceptionResNetV2 along with transfer learning techniques which helps to classify brain tumor histology image. Features extracted by InceptionResNetV2 are passed into an autoencoder network where non-linear transformation is applied to the extract features. Now generated new features are clustered by mean-shift clustering to give better result. 9083 histology image of 182 patients are taken from TCGA database. It is very hard to perform alone with deep learning techniques, so combined transfer learning with deep learning techniques. Transfer learning obtain basic parameters and train a deep learning model. Again it re-train fully connected layer for new target. This method obtains good result even on small datasets. Transfer learning uses the prior knowledge that is gained on one domain on another domain for classification and feature extraction. To have good performance, InceptionV3 and InceptionResNetV2 networks are adapted and it learn brain tumor histology image through transfer learning techniques. To solve unbalanced data problem, sample distribution of brain tumor histology images are expanded by rotation, inversion and with some other augmentation techniques. InceptionResNetV2 is mainly adapted to perform binary and multi-class classification of brain tumor histology image, also to obtain better performance than InceptionV3.

Autoencoder network is constructed to apply non-linear transformation and also to extract feature. Output of this autoencoder is passed to mean-shift clustering.

II. PROBLEM STATEMENT

Histopathology tissue analysis is completely a manual way of analysis and it is very hard to process due to large number of patient's data. It is a burden for pathologist to analyze and also generate time consuming problem. There are some cases where tissue diagnosing and surgery have to be done simultaneously and this give pressure to pathologist. Lack of publically available data and high quality labelled data is even rarer. Number of images (datasets) is necessary for training medical image. Data imbalance in the training set is also an issue in medical image analysis. Sometimes this number of images in the training data may be found with normal and non-pathological images. Rare diseases are missed without diagnosing due to adequate training examples. Glioblastoma image features derived only from H and E stained slides. Health results being studied by a non human actor is not trust worthy for detecting malignant or benign tumors.

III. DATASET

Datas are taken from TCGA database. 9083 images of 182 brain tumor patients are taken. To maintain original structure and molecular composition, each image taken through biopsy process. It is the process of taking tiny tissue by operating the tumor. These images are analysed by pathologist based on H and E stains. Finally images are classified based on their observation under microscope. Size of the image is 700*460 under RGB micrographs. Images of different magnification factors are defined as 40X, 100X, 200X, 400X. All these images are classified as benign or malignant tumors. Malignant types are Astrocytomas, Ependymomas, GBM-Glioblastoma Multiforme, Medulloblastomas and Oligodendroglioma. TCGA datasets are shown in table 1. Images are resized to 299*299. For some images pre-processing methods are done through Tensorflow framework by adjusting image size, reducing box border, saturation adjustment etc. Input size of image is based on RGB channel and based on pixel values of each channel that is normalized with the interval of [-1, 1]. To have good classification result, images of four magnification factors are given randomly to training and testing process with proportion of 7:3.

IV. METHODOLOGY USED

Transfer learning is a process that uses machine learning techniques, here the pre-trained model of certain task is re-used for another related task. Transfer learning approaches are Develop Model Approach and Pre-trained Model Approach. Transfer learning helps to diagnose disease efficiently. CNN networks take more time to converge, in finding learning rate and to normalize. Deep layer identify data based features, where this transfer learning learn basic features like edges, shapes, colors etc.

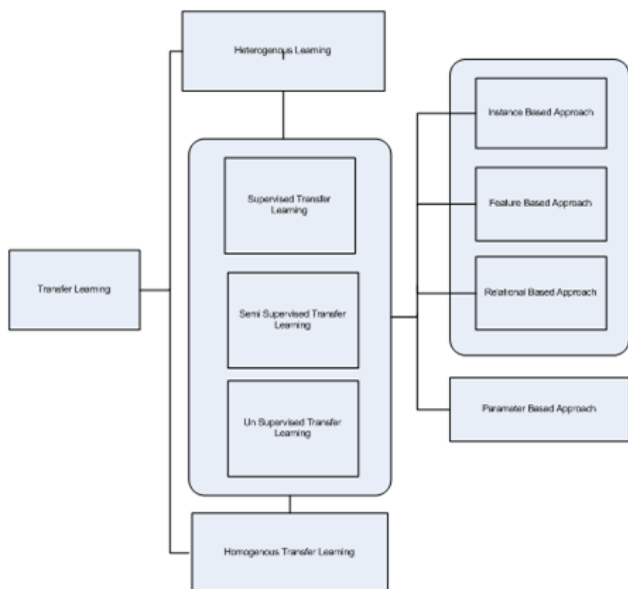


Fig. 1. Overall working of transfer learning.

V. RELATED THEORIES

Deep learning is currently popular in the field of computer vision and pattern recognition. Especially in computer based biomedical image analysis. One of the popular Deep Learning techniques is Convolution Neural Network (CNN), which is mostly applied in image classification. CNN is used in Glioma histopathology image to learn, extract features and to classify tumor. Image is given as input for CNN and produces probability values as output. It basically performs operations through hidden layers and extracts some high level features that can effectively represent the target classes. Its important operations are Convolution, Max Pooling, Rectified Linear Units (ReLU). High level features are extracted by performance of convolution and pooling. On the extracted features, classification is performed by fully connected layers. Extracted features are converted into vectors of huge parameters which need to be adjusted based on loss. This weight adjustment is done through back propagation algorithm.

A. Inception Network Classification Analysis

InceptionV3 and InceptionResNetV2 networks were developed by szegedy et al. in 2016 and 2017. This network was demonstrated in ILSVRC competition. InceptionV3 network achieve 78.0 percent accuracy in top1 error rate and 93.9 percent accuracy in top5 error rate. InceptionResNetV2 achieve 80.4 and 95.3 percent accuracy in top1 and top5 error rate. Important difference between both networks is InceptionResNetV2 have residual network. InceptionResNetV2 defeat InceptionV3 due to its residual connections which work more efficiently when applied to big data. Also work better on small datasets. InceptionV3 architecture shown in figure 2.

Table- II: Number of malignant tumors in different magnification. Astrocytomas(AC), Ependymomas(ED), GBM-Glioblastoma Multiforme, Medulloblastomas (MB) and Oligodendroglioma (OD).

Magnification	Malignant				
	AC	ED	GBM	MB	OD
40X	1253	543	456	254	189

100X	479	289	164	487	384
200X	386	241	170	632	408
400X	543	459	346	145	412

Two networks are similar with first layers as convolution and pooling. Middle layer is for inception modules that are stacked together. Output generated through fully-connected layer by using softmax function. InceptionV3 network consist of different filter with the size of 1*1, 1*3, 3*1 for network adaptability in different convolution kernel.

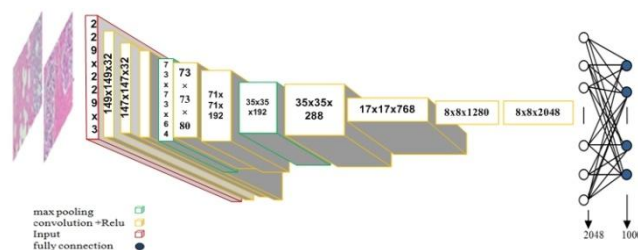


Fig. 2. InceptionV3 network.

InceptionResNetV2 have large number of layers, also residual units is added to each inception module to avoid occurrence of gradient problem. Figure 3 shows the structure of InceptionResNetV2.

B. Deep Feature Extraction

In other words transfer learning is said to be deep feature extraction. Deep layers extract high level features which help in image classification where CNN layers find basic features like edges, shapes etc.

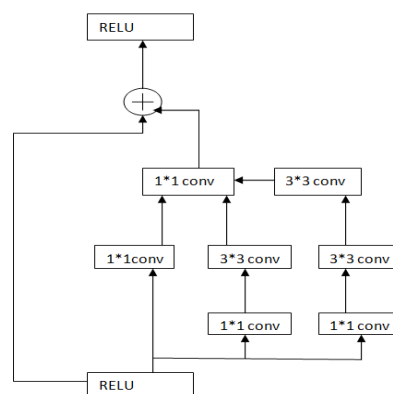


Fig. 3. Structure Of InceptionResNetV2 network.

C. Histology Based Image Analysis

Glioblastoma image features are derived from H and E stains. Malignant (cancer tumors) characteristics of the tumor can be identified based presence and absence of certain histological features, including mitotically active cells, nuclear atypia, micro vascular proliferation (enlarged blood vessels) and by the presence and absence of necrosis.

H and E stains are universally used for histology image examination. High cross validation with improved transfer learning can be implemented. Glioma histology image is given as input and passed into InceptionResNetV2 for feature extraction. Obtained features are classified by SVM classifier.

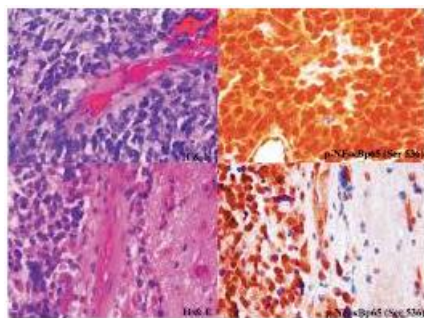


Fig . 4. Hematoxylin stains are acidic molecules, shades of blue. Eosin stains are basic materials shades of red, pink and orange.

VI. PROPOSED WORK

A. Transfer Learning Working On Glioma Histopathology

Transfer learning uses prior knowledge that is gained on one domain to another domain for classification and feature extraction [9]. Now pre-trained CNN is trained again (fine-tuning) on new datasets. Because of fine-tuning pre-trained CNN, it is widely used in deep learning applications, also seems to be much faster and easier than CNN model. At initial stage CNN learn edges, curves, corners like basic features and at final layer it extract specific features [8]. From pre-trained CNN, fully connected layer and softmax layer are discharged and the extracted features are classified by new classification scheme.

Here used InceptionV3 and InceptionResNetV2 network models, parameters are downloaded and trained on ImageNet dataset. This dataset composed of 1.2 million training images, 50,000 validation images and 100,000 testing images with 1000 different categories. In proposed work frozen all parameters of classification part and modified the number of neurons at fully-connected layer as per features extracted by InceptionResNetV2. 2 for binary classification and 5 for multi-class classification. Now the brain tumor histology images are trained with this fully-connected layer. Modified network is shown in figure 5.

Classification process is done on Tensorflow framework. Adam (adaptive moment estimation) algorithm is used in training process to perform optimization, with the iteration of 70 epochs. Batch size is set to 32 with initial learning rate of 0.0002. To reduce learning rate exponential decay method is used.

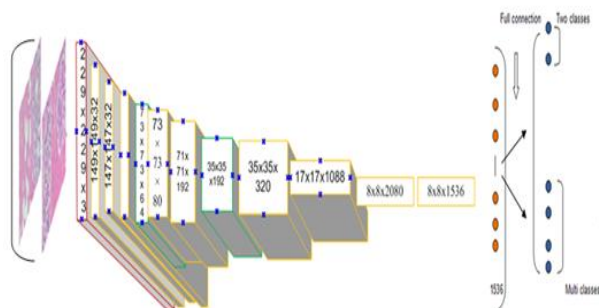


Fig. 5. InceptionResNetV2 architecture with 1536 vector for two class and multi class classification.

This is to ensure that the model moves through iteration quickly at its initial training stage. This helps to find stability at further iterations and helps to obtain quick optimal solution. Decay coefficient is set to 0.7 and decay speed is also fixed, so that decay occurs at every two epochs. Decayed learning rate is current learning rate. learning rate is initial learning rate, decay rate is decay coefficient, global step is current iteration step and decay step is decay speed.

$$\text{decayed learning rate} = \text{learning rate} \times \text{decay rate}^{\left(\frac{\text{global_step}}{\text{decay_steps}}\right)} \quad (1)$$

B. Evaluation Of Classification Result

To measure accuracy of classification, classification results are measured with some benchmarks. They are sensitivity (Se), specificity (Sp), positive predictive values (PPV), diagnostic odds ratio (DOR), F1 measure (F1), area under receiver operating characteristic curve (AUC), kappa criteria (kappa), Macro-F1, Micro-F1, image level test accuracy (ACCIL), patient level test accuracy (ACCPL). Macro-F1 and Micro F1 are the two variants of F1 for multi-class classification problems. Macro-F1 is average of F1 of each class. Micro-F1 is F1, this depends on precision and recall defined by sum of TP (true positive), FP (false positive), FN (false negative) of all classes. TP is number of malignant tumors correctly identified in testing subset. FP is number of malignant tumors incorrectly identified in testing subset. TN number of benign tumors correctly identified in testing subset. FN number of benign tumors incorrectly identified in testing subset. Se is the ratio of recognized malignant tumors from all malignant tumors in testing subset. Sp is the ratio of recognized benign tumors from all benign tumors images. Se and Sp are the accuracy of positive and negative class. PPV is the ratio of correctly identified malignant tumors from all recognized malignant images in the testing subset. Its precision is $F1 = \frac{2 * \text{precision} * \text{recall}}{(\text{precision} + \text{recall})}$ and $\text{recall} = \frac{TP}{(TP + FN)}$. DOR is the ratio of TP and TN to product of FP and FN. DOR change to infinity when related classifier is perfect. Diagnosis system is reliable if $Se \geq 80$ percent, $Sp \geq 95$ percent, $PPV \geq 95$ percent and $DOR \geq 100$ percent. Image level accuracy (ACCIL) by the ratio of Nrec (the number of brain tumor histology images correctly identified in testing subset), Nall (total number of brain tumor histology image in testing subset). Patient level test accuracy (ACCPL). It is the ratio of the sum of patients score to the total number of patients in the testing subset. Patient score is the ratio of Nrec to Np, that is, the ratio of correctly identified images of patients P from all images of patients P in the testing subset. Precision is same as PPV and recall is ratio of correctly identified malignant tumors to number of malignant tumor images in testing subset. AUC is the area under ROC curve, which is used for evaluating binary classification models. Values range from [0 to 1] represents better model performance. AUC is used by calling roc-auc-score function from scikit-learn library. It is python package (sklearn). In kappa where P0 is the image level test accuracy,

Pe is the ratio of sum of product to the number of real images in each category and predicted number of images in that category to the square of total samples. Calculation of kappa coefficient is based on confusion matrix. Kappa is used for consistency checking, its values ranges from [-1,1]. It can be divided into six groups of consistency levels as - 1~0.0 (poor), 0.0~0.20 (slight), 0.21~0.40 (fair), 0.41~0.60 (moderate), 0.61~0.80 (substantial), 0.81~1 (almost perfect).

C. Clustering Analysis

In previous section described classification analysis of brain tumor histology images. Supervised learning method needs experienced pathologist to analyse tumor and to assign labels for them as benign/malignant. It leads to time consuming problem. So to solve this clustering technique is adapted and this does not need any labels for samples. Only similarities are used to group them into different clusters. Therefore, samples present in a cluster are similar to each other, varies to those from other clusters.

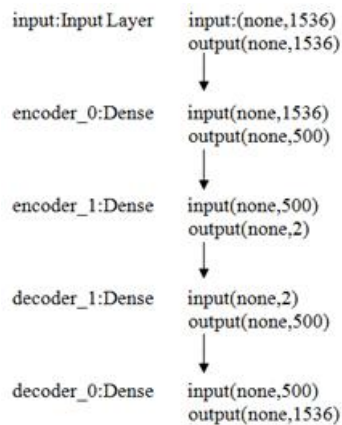


Fig. 6. Auto encoder.

InceptionResNetV2 is adapted to extract features and in that extracted features clustering analysis is performed. INRV2 network is mainly used for its efficient performance in classifying the images through automatic feature extraction. INRV2 output 1,536 dimension vector which act as input for clustering algorithm. Mean shift clustering is used to perform clustering.

Mean shift clustering is a sliding window based algorithm which helps to find dense areas of data points. It locate the center points of each group and then update the center points to be the mean of points within the sliding window, so it is called as centroid based algorithm. This algorithm start with circular shaped sliding window that is centered at a point with radius r as kernel. It is hill climbing algorithm that move kernels iteratively towards high density region until it converge. By moving to the mean of points in the window it automatically moves towards high density region. Now data points are clustered as per the sliding window in which they reside. As k mean clustering no need to choose number of clusters also cluster centers converge towards maximum density point. This algorithm is specially used because tumors are identified by high density area that differs from normal tissue. Based on the properties, value range and category described in Table 2, the tumors are identified.

Table- II: Above are properties that help to identify different features of tumors.

Properties	Value	Category
------------	-------	----------

	Range	
Clump Thickness	1 To 10	2-Benign 4-Malignant
Uniformity Of Cell Size	1 To 10	
Uniformity Of Cell Shape	1 To 10	
Marginal Adhesion	1 To 10	
Single Epithelial Cell Size	1 To 10	
Bare Nuclei	1 To 10	
Bland Chromatin	1 To 10	
Normal Nucleoli	1 To 10	
Mitoses	1 To 10	

Figure 6 displays the architecture. There are 2 encoded layers each with 500 neurons and 2 decoded layer to reconstruct the original image. By using autoencoder, 1,536 dimension feature vector extracted by InceptionResNetV2 network and it is transformed into two dimension feature vector as per training layer mentioned in figure 7. 2-dimension feature vector is set as input for median shift clustering to analyse brain tumor histology image.

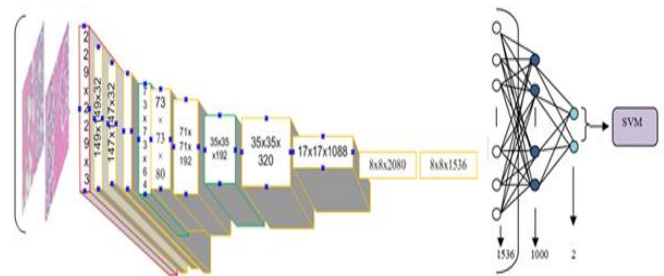


Fig. 7. InceptionResNetV2 network with auto encoder and SVM classifier.

VII. EXPERIMENTAL RESULTS

Classification result using InceptionV3 and InceptionResNetV2 network on brain tumor histology image is described here. Also comparison between raw dataset and augmented dataset is described. Experimental platform configuration is core i7-4810 CPU and 32 GB memory.

A. Experiment On Raw Data

To perform binary classification on brain tumor histopathological images, we used InceptionV3 and InceptionResNetV2 Networks. Table 3 describe experimental result using InceptionV3 and InceptionResNetV2 network. This performs binary classification in terms of Se, Sp, PPV, DOR, ACCIL, ACCPL, F1, AUC and kappa. Described InceptionV3 as INV3 and InceptionResNetV2 as INRV2. Multi-class classification will be useful in identifying exact stage of disease and helps to give proper treatment. Table 3 shows experimental results of brain tumor histology image using InceptionV3 and InceptionResNetV2 network. This performs multi-class classification in terms of ACCIL, ACCPL, Macro-F1, Micro-F1 and kappa. InceptionResNetV2 produce good result compared to InceptionV3.

Glioblastoma Multiforme Classification by Deep Learning Techniques on Histopathology Images

Residual connection of InceptionResNetV2 avoid vanishing gradient problem that occurs due to more number of layers. It also extracts informative features and increases network performance. Result on magnification factor 40X is good while comparing with other magnification factors. Se \geq 98 percent, Sp \geq 92 percent, PPV \geq 96 percent and DOR \geq 100 on InceptionResNetV2. At magnification factor of 40X Se \geq 98 percent, Sp \geq 96 percent, PPV \geq 98 percent and DOR \geq 100. Values of AUC and kappa from table 3 shows best binary classification. Value of kappa from table 4 give perfect multi-class classification result. InceptionResNetV2 network on multi-class classification produces best classification result except on 400X magnification factor, but it achieve moderate result.

Table- III: Evaluation of binary classification using InceptionV3, InceptionResNetV2 in percentage.

Classification	Network	Criteria	Magnification Factor			
			40X	100X	200X	400X
Binary	INV3	Se	98.02	98.47	99.02	96.43
		Sp	94.33	93.44	91.42	90.98
		PPV	97.40	96.65	95.87	95.86
		DOR	81,230	92,303	106,701	27,103
		ACC_IL	96.82	96.74	96.47	94.70
		ACC_PL	97.73	94.20	87.19	96.66
		F1	97.70	97.55	97.41	96.14
		AUC	99.46	99.03	99.29	97.91
		Kappa	92.66	92.76	91.94	87.69
		IRV2	Se	98.45	98.89	99.12
	Sp		96.64	92.95	92.81	92.09
	PPV		98.46	96.43	96.40	96.52
	DOR		185,774	118,782	147,138	58,835
	ACC_IL		97.88	96.88	96.97	96.96
	ACC_PL		98.02	97.05	82.73	88.11
	F1		98.48	97.65	97.73	97.26
	AUC		99.58	98.85	99.60	98.80
	Kappa		95.13	92.96	93.19	91.06

Table- IV: Evaluation of multi-class classification using InceptionV3, InceptionResNetV2.

Classification	Network	Criteria	Magnification Factor			
			40X	100X	200X	400X
Multi-class	INV3	ACC_IL	90.27	85.34	83.99	82.09
		ACC_PL	90.45	89.04	80.64	81.09
		Macro-F1	88.55	82.57	79.65	77.97
		Micro-F1	90.29	85.34	83.99	82.08
		kappa	87.36	80.27	77.91	76.39
	IRV2	ACC_IL	92.08	88.07	87.62	84.51
		ACC_PL	89.12	88.45	86.08	71.42
		Macro-F1	90.88	85.66	84.09	80.14
		Micro-F1	92.08	88.06	87.63	84.50
		kappa	89.74	84.04	82.85	79.70

Power of this approach is calculated by p-values of AUC and kappa. It's the probability that measure statistical evidence against null hypothesis. Low p-values reject null hypothesis. Calculated p-value of AUC and kappa, then compared p-value to its significance level of α . α and it is set to 0.05. Considered p-value of AUC and kappa in both binary and multi-class classification using InceptionResNetV2. AUC and kappa p-values are calculated in equation 2 to 5 and pnorm function is performed in R.

$$SE_{AUC} = \frac{\sqrt{0.25 + (na + nn - 2)}}{na \times nn \times 12} \quad (2)$$

$$Z_{AUC} = \frac{A - 0.5}{SE_{AUC}} \quad (3)$$

$$SE_{kappa} = \frac{\sqrt{p_0 \times (1 - p_0)}}{\sqrt{N} \times (1 - p_e)} \quad (4)$$

$$Z_{kappa} = \frac{kappa}{SE_{kappa}} \quad (5)$$

Above na and nn in (2) is the number of malignant and benign tumor in testing subset. A in 3 is the value of AUC. N in (4) is the total number of samples. kappa in (5) and Z value to AUC in (3) are converted into p-value using pnorm function of R. Except binary classification p-values of AUC is p=6.88e-85(40X), p=2.24e-89(100X), p=3.73e-89(200X), p=9.20e-75(400X). p-values for all kappa are 0.0, p-value for AUC and kappa is \leq 0.05. This helps to reject null hypothesis, which are random guesses and helps to accept static predictions.

Experiment On Augmented Datasets

By comparing binary and multi-class classification, multi-class classification gives least result, so that the confusion matrix of multi-class classification is processed with further more analysis. Through observation, it is found that benign tumors are identified as malignant tumors. This leads to high false positive rate. Due to data unbalance incorrect classification occurs. Extracted features were not able to identify malignant subclasses which are less in number of samples. Subclasses with less samples are incorrectly classified as samples which are many in numbers. To get rid of this misclassification and to avoid false positive the original samples are expanded. Ac (astrocyte) sample is taken as baseline for all magnification factor and amplified each other remaining subclasses by turning left, right, up, down and towards various degrees. Extended datasets are randomly partitioned for training and testing phase in the ratio of 7:3. By using transfer learning technique, INRV2 is re-trained to get better diagnosis. To evaluate the difference in loss function between raw and augmented dataset during training process, we changed the values of loss function with variations in number of epochs. Compared loss function on 40X magnification factor using INRV2 network and observed the changes.

Figure 8 shows the evaluation result of loss function in both binary and multi-class classification using INRV2 network on raw and augmented datasets. For binary and multi-class classification, deep learning parameters do not change. Table 5 and 6 shows the evaluation result on raw and augmented datasets in binary and multi-class classification. Beyond raw data, augment data converge short and smooth to small values which lead to obtain less loss rate.

This is same for both binary and multi-class classification.

Table- V: Evaluation of binary classification on raw data and augmented data using InceptionV3, InceptionResNetV2.

Classification	Network	Criteria	Magnification Factor			
			40X	100X	200X	400X
Binary	Raw Data	Se	98.47	98.88	99.12	98.07
		Sp	96.64	92.94	92.79	92.11
		PPV	98.45	96.46	96.36	96.52
		DOR	185,774	118,782	147,135	58,836
		ACC_IL	97.89	96.90	96.97	96.97
		ACC_PL	98.02	97.06	82.75	88.12
		F1	98.46	97.65	97.75	97.26
		AUC	99.56	98.83	99.61	98.81
		Kappa	95.11	92.97	93.18	91.05
	Aug Data	Se	99.96	99.43	99.64	98.86
		Sp	99.63	99.26	99.19	99.33
		PPV	99.66	99.38	99.31	99.42
		DOR	56122,884	2441,732	3427,111	1324,245
		ACC_IL	99.79	99.36	99.42	99.11
		ACC_PL	99.94	99.96	99.99	99.91
		F1	99.82	99.42	99.48	99.14
		AUC	100.0	99.99	99.96	99.98
		Kappa	99.59	98.77	98.89	98.19

Table- VI: Evaluation of multi-class classification on raw data using InceptionV3, InceptionResNetV2.

Classification	Network	Criteria	Magnification Factor			
			40X	100x	200X	400X
Multi-class	Raw Data	ACCIL	92.08	88.04	87.63	84.50
		ACCPL	89.10	88.46	86.08	71.43
		MacroF1	90.88	85.66	84.09	80.12
		MicroF1	92.08	88.05	87.64	84.51
		kappa	89.74	84.03	82.83	79.70
	Aug data	ACCIL	97.63	97.00	96.90	97.50
		ACCPL	98.43	98.08	97.86	97.41
		MacroF1	97.68	97.05	97.03	97.48
		MicroF1	97.64	97.00	96.90	97.50
		kappa	97.29	96.54	96.46	97.13

Table 4 and 5 describes that augmented data perform perfect than raw data. At 40X magnification factor, obtained result is better than other magnification factors. $Se \geq 98$ percent, $Sp \geq 92$ percent, $PPV \geq 96$ percent and $DOR \geq 100$ in binary classification, it is performed on all magnification factors of raw and augment dataset shown in Table 5. At augment dataset obtained result is $Se \geq 98$ percent, $Sp \geq 99$ percent, $PPV \geq 99$ percent and $DOR \geq 100$.

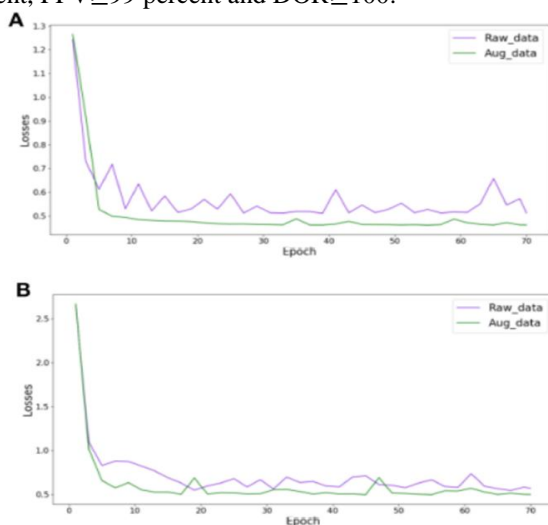


Fig. 8. Change in loss function while training IRV2 network on raw and augment data. A) Binary. B) Multi-class classification.

Augment data on INRV2 network, produces perfect result on both binary and multi-class classification. AUC value is 1.0 on augment dataset at 40X magnification factor, this shows that binary classification performs good. kappa values at Table 5 and 6 describe that best result is achieved on augment dataset for multi-class classification. AUC and kappa, p-value is 0.0 which is ≤ 0.05 , this says that null hypothesis will be rejected (random prediction) and accept static significance predictions.

VIII. CONCLUSION

Proposed system, with InceptionV3 and InceptionResNetV2 network used transfer learning techniques to analyse brain tumor histology image. Structure and parameters learned are frozen and then the numbers of neurons on fully connected layers are changed as per proposed task. These parameters are re-trained for binary and multi-class classification. InceptionResNetV2 network perform good when compared to InceptionV3 network. Best results are given on augment datasets when compared to raw datasets, particularly it works well on multi-class classification. INRV2 network extract most needed and informative features when comparing with other network. Autoencoder helps to detect an effective and low dimension feature that is present in brain tumor histology image. SVM is a good classifier that helps to classify the data that comes from autoencoder. Variations in image resolution and the images from same class are similar with different classes and the image contrast which causes difficulties in classification are some of the problems which will be focused in future.

REFERENCES

1. H.O. Lyon, A.P. De Leenheer, R.W. Horobin, W.E. Lambert, E.K.W. Schulte,
2. Van Liedekerke, D.H.Wittekind, "Standardization of reagents and methods used in cytological and histological practice with emphasis on dyes, stains and chromogenic reagents, Histochem". J. 26 (7) (1994) 533 - 544.
3. K. Tamaki, K. Fukuma, H. Kawanaka, H. Takase, S. Tsuruoka, B. J. Aronow, and S. Chaganti. "Comparative study on feature descriptors for brain image analysis", International Conference on and Advanced Intelligent Systems (ISIS), Proc. IEEE, pp. 679 - 682, 2014.
4. 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 Knowledgebase and Intelligent Information and Engineering Systems (KES), pp. 1202 - 1210, 2016.
5. K. Fukuma, H. Kawanaka, V. B. S. Prasath, B. J. Aronow, and H. Takase, "A study on feature extraction and disease stage classification for glioma pathology images", IEEE International Conference on Fuzzy Systems (FUZZ - IEEE), 2016.
6. A. Yonekura, H. Kawanaka, V. B. S. Prasath, B. J. Aronow, and H. Takase,
7. "Glioblastoma Multiforme Tissue Histopathology Images Based Disease Stage Classification with Deep CNN", in the 6th International Conference on Informatics, Electronics & Vision (ICIEV), 2017
8. Yonekura A, Kawanaka H, Prasath VBS, Aronow BJ, Takase H, "Improving the generalization of disease stage classification with deep CNN for glioma histopathological images", International workshop on deep learning in bioinformatics, biomedicine, and healthcare informatics (DLB2H); 2017. pp 1222 - 1226.

9. Asami Yonekura, Hiroharu Kawanaka, V. B. Surya Prasath, Bruce J. Aronow, Haruhiko Takase, "Automatic disease stage classification of glioblastoma multiforme histopathological images using deep convolutional neural network", Korean Society of Medical and Biological Engineering and Springer-Verlag GmbH Germany, part of Springer Nature 2018
10. Krizhevsky A, Sutskever I, Hinton GE, "Imagenet classification with deepconvolutional neural networks", In: Advances in neural information processing systems, pp. 1106 - 1114. 2012.
11. Orenstein EC, Beijbom O. "Transfer learning and deep feature extraction for planktonic image data sets", In: 2017 IEEE Winter conference on applications of computer vision (WACV). IEEE. 2017.
12. Krizhevsky A, Sutskever I, Hinton GE. "Imagenet classification with deepconvolutional neural networks". In: Advances in neural information processing systems, pp. 1106-1114. 2012.
13. Russakovsky O, et al. "Imagenet large scale visual recognition challenge". Int J Comput Vis. 2015;115(3):211-252.
14. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. "Support vector machines". In: IEEE intelligent systems and their applications, vol 13(4), pp.18-28. July-Aug 1998.
15. Vedaldi Andrea, Zisserman Andrew. "Efficient additive kernels via explicit feature maps". IEEE Trans Pattern Anal Mach Intell. 2012; 34(3):480-92.
16. P. Sobana sumi, Radhakrishnan Delhibabu. "Glioma Multiforme on High Resolution Histology Image Using Deep Spatial Fusion Network". EEML 2019, to be published.

AUTHOR PROFILE



P. Sobana Sumi received the B.Tech. degree in Information Technology from the National College of Engineering, Maruthakulam, Anna University, Tirunelveli, India, in 2011. Also received M.E. degree in Computer Science and Engineering from SCAD College Of Engineering And Technology, Cheranmadevi, Anna University, Chennai, India, in 2015. Currently doing research in medical

images with deep learning, at the school of computer science and engineering, Vellore Institute OF Technology, Vellore, India. Published paper in the title of "Glioblastoma Multiforme Classification on High Resolution Histology Image Using Deep Spatial Fusion Network" at EEML workshop 2019 (to be published), Perm, Russia. Her research interest includes Image processing, Deep Learning, Machine Learning, Cloud computing.



Radhakrishnan Delhibabu received B.Sc degree in Computer Science from Madras University, Tamilnadu, India, in 2002. Received his M.Sc degree in Computer Science from Madras University, Tamilnadu, India, in 2004. Received his M.Phil degree in Computer Science from Bharathidasan University, Tamilnadu, India, in

2005. Received his M.E. degree in Computer Science and Engineering from Vinayaga Mission, Tamilnadu, India, in 2009. Received his Doctorate, Ph.D in Computer Science from Anna University Chennai, India and RWTH Aachen, Germany, Research Fellowship sponsored by SSN College of Engineering and Erasmus Mundus fellow by India4EU, in 20014. Currently working as Associate Professor at School of Computer Science & Engineering (SCOPE), VIT University, Vellore, India. Also doing D.Sc in Mathematical Logic, Kazan Federal University, Kazan, Russia. Received Best Paper Award from AIST 2014 at Yekaterinburg, Russia, Best PhD Thesis from KI 2015 at Dresden, Germany, Best Invited Speaker Award from AIST 2016 at Yekaterinburg, Russia. Honoured by Indian International Friendship Society for Marquis Who's Who, with Rashtriya Gaurav Award. Published many paper in various conferences, journals, workshops. Wrote several book chapters also. His research interests are Abductive Inference, Belief Revision, Knowledge Base System, Ontologies, Description Logic, Cognitive Robotics, Intelligent Agents, Decision Support Systems, Knowledge Representation and Reasoning, Reasoning under Uncertainty, Real Life Application, Constraint Satisfaction, Datalog, Prolog, Database Application, DDB, Addiction, Artificial Consciousness, Affective Computing, Decision Making, Meditation and Time Perception, Deep learning, Data Mining, Formal Concept Analysis, Artificial Intelligence, Information Retrieval, Big Data and Machine Learning, Optimal Recovery Schemes, A Scalable, Fault Tolerant, Heterogeneous, Distributed Algorithms, Secured Distributed Storage Framework.