Machine Learning and Deep Learning Techniques to Predict Overall Survival of Brain Tumor Patients using MRI Images

Lina Chato
Department of Electrical and Computer Engineering
University of Nevada, Las Vegas (UNLV)
Las Vegas, Nevada
Lina.chato@unlv.edu

Shahram Latifi
Department of Electrical and Computer Engineering
University of Nevada, Las Vegas (UNLV)
Las Vegas, Nevada
Shahram.latifi@unlv.edu

***Abstract*—This paper presents a method to automatically predict the survival rate of patients with a glioma brain tumor by classifying the patient's MRI image using machine learning (ML) methods. The dataset used in this study is BraTS 2017, which provides 163 samples; each sample has four sequences of MRI brain images, the overall survival time in days, and the patient's age. The dataset is labeled into three classes of survivors: short-term, mid-term, and long-term. To improve the prediction results, various types of features were extracted and trained by various ML methods. Features considered included volumetric, statistical and intensity texture, histograms and deep features; ML techniques employed included support vector machine (SVM), k-nearest neighbors (KNN), linear discriminant, tree, ensemble and logistic regression. The best prediction accuracy based on classification is achieved by using deep learning features extracted by a pre-trained convolutional neural network (CNN) and was trained by a linear discriminant.**

Keywords—Convolutional neural network, statistical texture features, intensity texture features, volumetric features, deep learning features, Support vector machine, glioma, logistic regression, pre-trained CNN, discrete wavelet transform, K-nearest neighbors, gray level co-occurrence matrix.

I. INTRODUCTION

A glioma, also known as a high-grade glioma (HGG), represents those tumors arising from the gluey or supportive tissue of the brain [1]. This type of tumor is thought to be the most aggressive type of a brain tumor, and overall survival does not exceed two years [2, 3]. Based on statistics from the American Brain Tumor Association (ABTA) [1], gliomas represent 74.6% of all malignant tumors and 24.7% of all primary brain tumors. Survival after diagnosis with a primary brain tumor depends on such significant factors as age, histology, molecular markers, and tumor behavior [1].

The multimodal brain tumor segmentation (BraTS) challenge focuses on the evaluation of state-of-the-art methods for the segmentation of brain tumors in magnetic resonance imaging (MRI) scans. For BraTS 2017 challenge [4, 5], the BraTS 2017 challenge released two tasks based on MRI modalities: 1) a segmentation task and 2) a prediction task.

Recently in the medical field, machine learning (ML) and deep learning techniques have been used widely in automatic detection, classification, and regression [6-19]. In particular, ML methods are accurate tools for small datasets when the meaningful features are extracted for training a model. However, deep learning renders high accuracy without the need to use any specific meaningful features to train a model, but only for datasets that are very large. Texture features provide information that describes the arrangement of color or intensities in an image. Therefore, it is used widely for cancer detection and classification based on medical images. Gardezi and Faye [6] used a circular local binary pattern (CLBP) and curvelet texture features that were extracted from thermogram images and combined to produce a feature fusion matrix. This matrix was used to train nearest neighbor classifier to detect breast cancer. This method achieved the detection accuracy 96%. In addition, Pramanik et al. [7] presented an automatic method to detect early breast cancer based on statistical texture features in a thermogram. The authors first applied Otsu's thresholding to define the region of interest (ROI), which was processed later by discrete wavelet transform (DWT). Then, the statistical texture features were extracted from the processed ROI and used to train the feed-forward artificial neural network with a gradient decent training rule. The automatic detection accuracy of this method is 90.48%.

Sachdeva and et al. [8] proposed a multi-class brain tumor classification based on statistical texture and intensity features trained by artificial neural network (ANN) and based on T1-weighted MRI images. The accuracy of this method improved from 77% to 91% by reducing the dimensionality of the features vector, using principal component analysis (PCA). Another study [9] focused on a method for brain tumor detection and classification based on extracting the shape, texture, and intensity features from ROI, segmented by multiple kernel probabilistic clustering (MKPC). The extracted features were optimized using linear discriminant analysis (LDA), and then trained by a deep neural network. Faro et al. [10] proposed an accurate method for diagnosing multiple sclerosis patients based on T2 MRI images, with a 95% accuracy. In this method, the texture features were based on gray level co-occurrence and the statistical moments of the gray-level histogram were extracted and trained by multi-layer perceptron (MLP) neural networks.

This work is supported in part by Graduate College at UNLV and in part by the Graduate and Professional Student Association at UNLV.

2471-7819/17/31.00 ©2017 IEEE
DOI 10.1109/BIBE.2017.00009

Although training deep learning involves a high computational cost and a large dataset, a convolutional neural network (CNN) has been used widely for medical image segmentation and classification [11-17], and has achieved high accuracy. In general, deep learning approaches focus on building a good deep structure instead of determining specific sophisticated features to describe a model efficiently.

To improve the detection and segmentation tasks in the medical field, various deep learning approaches have been proposed based on medical images. Paul et al. [11] used pre-trained CNN for deep feature extraction combined with traditional image features to improve the survival prediction accuracy in lung cancer computed tomography (CT) images. In another study, Paul et al. [12] improved the prediction results from their previous study [11] by using features from post rectified linear units (post ReLU) instead of the raw extracted features from the layer before the last, which is the fully connected layer (output layer).

To avoid overfitting and to capture three-dimensional (3D) information, Saifeng Liu et al. [13] augmented the data by means of 3D rotation and slicing, and presented a new deep learning structure, known as XmasNet, for the diagnosis of prostate cancer. The 3D deep feature extraction of this approach proved its superiority over two-dimensional feature extraction [14]. Nie et al. [14] proposed using 3D deep-features extraction to improve the classification of the survival time for brain tumor patients into two classes, (short-overall and long-overall survivors) with a threshold time of 22 months. Due to the various sizes, shapes, and locations of the brain tumors, Zhao and Jia [15] proposed an automatic brain tumor segmentation method based on multi-scale CNN. The authors found that their proposed method improved the results of tumor segmentation compared to using a single-scale input patch. For deeper structures, using small kernels were proposed by Pereira et al. [16] to improve detection and segmentation of malignant gliomas in MRI images of various modalities. Kumar et al. [17] proposed two cascaded CNNs to predict the recurrence of prostate cancer recurrence using tissue images; the first one was used to detect individual nuclei and the second was used to classify the nuclei.

The study described in this paper developed an accurate model to predict the overall survival of patients based on classification using the BraTS dataset for 2017 challenge. This paper is organized as follows. Section II presents the implementation of the overall survival (OS) prediction system and its components. Section III displays and discusses the experimental results. Finally, Section IV concludes this paper, and presents further suggestions to improve the classification in future research.

II. Overall-Survivor Prediction System

A. System Outline

Due to the small size of the data set provided [4,5], various ML methods were used to develop the proper overall survivor prediction based on the classification approach. The general structure of the proposed system consisted of three steps, as shown in Fig. 1. The first step represents the extraction of the meaningful features to describe each specific class from the labeled MRI-image modalities. Second step involved producing the prediction model by using the extracted features to training various ML methods. Finally, the last step involved evaluating the performance of each classifier by calculating the accuracy of each class separately as well as for the overall system.

B. Dataset

The glioma brain dataset was provided by the BraTS 2017 Challenge [4, 5]. This dataset contains two groups, a glioblastoma known as higher grade glioma (GBM/HGG), and lower grade glioma (LGG). Four sequences of MRI modalities were provided for each patient, with an additional file that contains the ground truth labels for the regions have a glioma brain tumor. The MRI modality scans were for:

- Native (T1),
- Post-contrast T1-weighted (T1Gd),
- T2-weighted (T2), and
- T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes.

Each MRI modality and segmentation labels file consisted of 155 slices in that are 240 x 240 pixels in size. The provided labels in the segmentation file are:

- Label 1 for the necrotic and non-enhancing tumor (NCR/NET),
- Label 2 for the edema tumor (ED),
- Label 4 for GD-enhancing tumor, and
- Label 0 for everything else.

Some of the labeled data was manually annotated by clinical experts in BraTS team, and the others were obtained from the highly ranked segmentation methods from previous BraTS challenges. The OS data was provided in .csv file format. This dataset consists of only the HGG patients, because that type was the malignant brain tumor that caused death for most brain tumor patients. This dataset contains the identification (ID) of the each patient, which includes: 1) the name of the folder that refers to the four MRI modalities and segmentation labels file, 2) the age of each patient, and 3) the OS of each patient in terms of days.

The BraTS 2017 challenge suggested predicting the OS either based on the regression or the classification tasks. For prediction based on the classification task, it was (is) suggested to divide the provided dataset into three groups (classes), based on the survival time: 1) short-term survivors for less than 10 months, 2) mid-term survivors for a range of 10-15 months, and 3) long-term survivors for more than 15 months.

C. Feature Extraction Methods

To develop accurate prediction-model-based classification for HGG brain glioma, various types of features are extracted.

The accuracy of a prediction model depends on various factors, for example, the size of the dataset, valuable samples, the type of extracted features, the number of features, the type of ML, and the parameters of an ML. Therefore, various ML methods – such as a support vector machine (SVM), K-nearest neighbors (KNN), linear discriminant, tree, ensemble, and logistic regression – are used to develop the proper prediction model based on classification. Volumetric and location features, statistical and intensity texture features, histogram distribution, and 2D deep features are used to train the above ML methods for improving prediction accuracy.

The feature extraction methods used in this paper are described as follows:

i. Volumetric and location features were extracted from each sample of the prediction BraTS dataset. The segmentation labels file was used to locate the regions of the tumor for extracting these types of features. The location features describe the distances and direction of the centroid of the whole tumor to the centroid of a brain for a specific sample. The centroid of both the tumor and the brain was calculated by using the K-means nearest method. The eight volumetric and location features were extracted for each sample (i.e., patient) to get the features vectors, which were: {distance, X, Y, Z, volumetric ratio of total tumor to the brain, volumetric ratio of NCR/NET tumor to the total tumor, volumetric ratio of ED tumor to the total tumor, volumetric ratio of the gadolinium (GD)-enhancing tumor to the total tumor. Figure 2 shows the three tumor regions in a slice from a specific sample from the dataset.

ii. Statistical and intensity texture features were based on DWT. In medical images, the gray level co-occurrence matrix (GLCM) often has been used to extract statistical texture features for classification and detection tasks [7, 8, 10, and 18] due to the high accuracy that could be achieved. GLCM provides certain information about the spatial distribution of the gray level in the texture image, although this information does not include the shape features. For increasing the meaning of these statistical features, DWT was used with GLCM by Hu et al. [18], in which improved the detection of osteosarcoma bone tumor in computed radiography (CR) images. In this paper, the db4 Level 3 wavelet transform was used to extract the wavelet features; then, PCA was applied to reduce the dimensionality of the wavelet features. GLCM properties were calculated from the reduced wavelet features, and then the statistical texture features based on DWT were obtained. The DWT provided time and frequency information of a signal. It decomposed the signal into detailed information from a high-pass filter, and approximation information from a low-pass filter. Two-dimensional DWT was used for image applications that decomposed the image into four components: approximation (cA), Vertical (cV), horizontal (cH), and diagonal (cD). The information from a low-pass filter (cA) was very important because it provided an accurate and reduced version of an image. The intensity features were calculated from the cA component of the image. The steps of extracting the statistical and intensity texture features are shown in Fig. 3. Thirteen overall texture features were extracted in this study; four statistical features were derived from GLCM, and nine intensity features were from cA3 of the db4 wavelet transform.

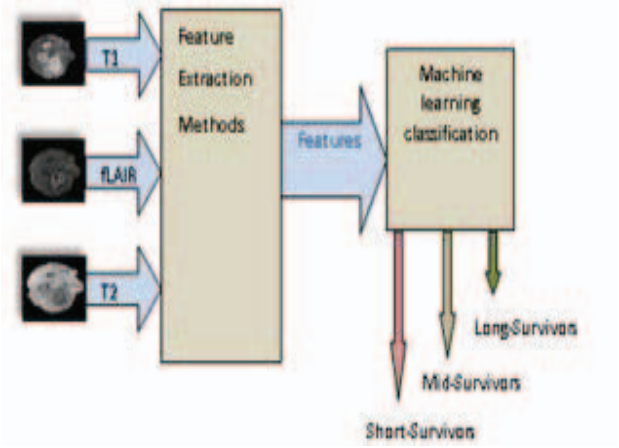

Fig. 1 Overall-Survival Prediction System.

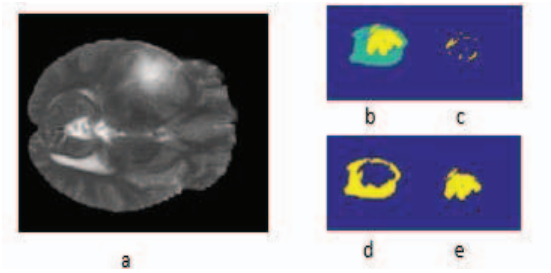

Fig. 2 Sample of an MRI image from the BraTS dataset: a) slice of T1-enhanced MRI, b) the three tumor labels, c) NCR/NET region, d) edema region, and e) and enhanced tumor.

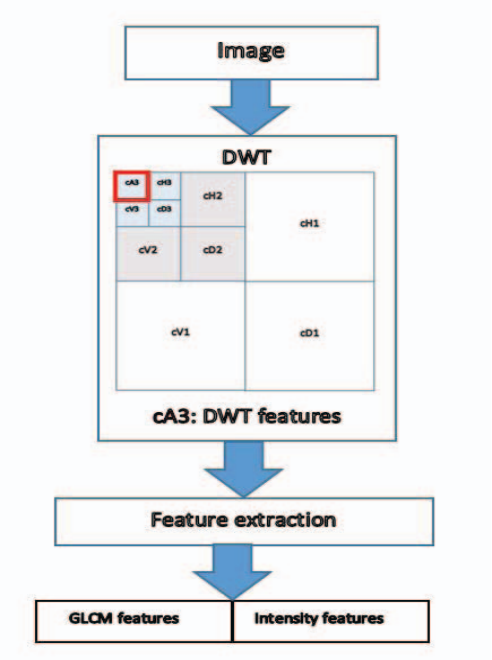

Fig. 3 Steps of extraction texture features.

The statistical texture features calculated from the GLCM are defined as follows.

Contrast defines the intensity contrast between a pixel and its neighbor over the whole image.

$$\boldsymbol{Contrast} = \sum_i \sum_j (\boldsymbol{i-j})^2 \boldsymbol{N_d(i,j)}\,, \tag{1}$$

where $N_d(i,j)$is the GLCM data of an image at location (i,j).

Correlation defines the spatial dependencies between a pixel and its neighbor over ROI in an image:

$$\boldsymbol{Correlation} = \sum_i \sum_j \frac{(i-\mu_i)(j-\mu_j)N_d\,(i,j)}{\sigma_i \sigma_j}\,, \tag{2}$$

where $\mu_i, \mu_j \; and \;\; \sigma_i, \sigma_j$ are the horizontal and vertical means and variances.

Homogeneity defines the closeness of the distribution of elements in the GLCM to the GLCM diagonal:

$$\boldsymbol{Homogeneity} = \sum_i \sum_j \frac{\boldsymbol{N_d\,(i,j)}}{\boldsymbol{1+|i-j|}} \tag{3}$$

Energy measures the similarity or uniformity of an image, and it is defined as angular second moment as follows:

$$\boldsymbol{Energy} = \sum_i \sum_j \boldsymbol{N_d^2(i,j)} \tag{4}$$

The intensity texture features were extracted from DWT features of an image, and they are defined as follows.

Mean (M) is the sum of the intensity values of pixels divided by the number of pixels in the ROI of an image:

$$\boldsymbol{M} = \frac{\boldsymbol{1}}{\boldsymbol{m*n}} \sum_{\boldsymbol{i=0}}^{\boldsymbol{m-1}} \sum_{\boldsymbol{j=0}}^{\boldsymbol{n-1}} \boldsymbol{f(i,j)} \tag{5}$$

where f is cA3 of an image with size (m x n) pixels, and $f(i,j)$is the value of a pixel at location (i,j).

Standard Deviation (SD) defines the spread of the gray level around the mean. It is the second central moment describing probability distribution of an image:

$$\boldsymbol{SD} = \sqrt{\frac{\boldsymbol{1}}{\boldsymbol{m*n}} \sum_{\boldsymbol{i=0}}^{\boldsymbol{m-1}} \sum_{\boldsymbol{j=0}}^{\boldsymbol{n-1}} (\boldsymbol{f(i,j)-M})^2} \tag{6}$$

Variance ($\boldsymbol{Var}$) is the square root of standard deviation:

$$\boldsymbol{Var} = \sqrt{\boldsymbol{SD}} \tag{7}$$

Skewness (S_k) is the measure of symmetry of the distribution. The skewness of a random variable X is denoted as $S_k(X)$, and is defined as:

$$\boldsymbol{S_k(X)} = \left(\frac{\boldsymbol{1}}{\boldsymbol{m*n}}\right) \frac{\sum(\boldsymbol{f(i,j)-M})^3}{\boldsymbol{SD^3}} \tag{8}$$

Kurtosis ($K_{urt)}$ is a measure of peakedness or flatness of the distribution. The forth central moment gives kurtosis. For the random variable X, the Kurtosis is denoted as $K_{urt}(X,)$and is defined as:

$$\boldsymbol{K_{urt}(X)} = \left(\frac{\boldsymbol{1}}{\boldsymbol{m*n}}\right) \frac{\sum(\boldsymbol{f(i,j)-M})^4}{\boldsymbol{SD^4}} \tag{9}$$

Entropy represents the amount of randomness in gray level intensity distribution of an image:

$$\boldsymbol{Entropy} = -\sum_{\boldsymbol{i}} \sum_{\boldsymbol{j}} \boldsymbol{f(i,j) log_2 f(i,j)} \tag{10}$$

Smoothness is a measure of grey level contrast that can be used to establish descriptors of relative smoothness:

$$\boldsymbol{Smoothness} = \boldsymbol{1} - \frac{\boldsymbol{1}}{\boldsymbol{1+SD}} \tag{11}$$

Inverse Difference Moment *(IDM)* is expressed as:

$$\boldsymbol{IDM} = \sum_{\boldsymbol{i}} \sum_{\boldsymbol{j}} \frac{\boldsymbol{f\,(i,j)}}{\boldsymbol{1+|i-j|}} \tag{12}$$

Root Mean Square *(RMS)* is expressed as:

$$\boldsymbol{RMS} = \sqrt{\frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \left|\boldsymbol{f_{i,j}}\right|^2}{\boldsymbol{m}}} \tag{13}$$

iii. Histogram. The MRI modality is used to extract the histogram features. For extracting the histogram for each slice, the ROI was located as the overall brain region. This extraction avoided the high density rate of the black pixel in the background of the MRI images. For extracting the ROI in each slice, Otsu's thresholding method was applied. The histogram features for each sample were calculated by combining the histogram of the 155 slices. Figure 4 shows the histogram for Samples 67 to 72 in the dataset.

iv. Deep feature extraction. T: transfer training based on a pre-trained CNN approach was used to extract 2D global and local features from the ROI. The pre-trained deep CNN was used when it was difficult to decide the type of valuable features needed to develop an accurate model. Furthermore, only a small training dataset was provided. Three pre-trained deep CNN were provided in MATLAB 2017: AlexNet, VGG16, and VGG19. The pre-trained CNN AlexNet, was used for feature extraction, is a deep CNN for image classification that won the ILSVRC-2012 competition, and achieved a winning top-5 test error rate. AlexNet has eight layers [19]; the first five are convolutional and the last three are fully connected layers, with maximum pooling layers used in between the convolution and fully connected layers. The ROI (i.e., the tumor) was extracted from each slice and fed to the AlexNet for feature extraction. The medical image with the ROI was resized to 224 x 224 x 3 to be compatible with the input of the AlexNet. For the z dimension (3 channels), the three modalities of MRI (T1-cd, T2, T2-flair) images were used. The size of the features vector was 4096, which was extracted from each image for training, using ML.

III. Experimental Results and Evaluation Performance

First, the volumetric and location features were extracted from each segmentation file provided in the dataset. The location features were calculated from volumetric features. The volumetric, location features, and the age were combined to create a features vector. The combined features vectors with the labels were used to train various ML. The best results obtained are summarized in Table 1. From many trails, the observation showed that the worst classification accuracy which caused reduction of the overall system classification accuracy was due to the bad classification accuracy of class 2, as the confusion matrix shows in Fig. 5. The accuracy test parameters "T", "P", "TPr", and "FNr" refer to "true class", "predicted class", "true positive rate", and "false negative rate" respectively. In [9], good classification results were obtained based on two classes: short-overall survivor and long-overall survivor with a threshold time of 22 months. Therefore, an experiment was conducted to test the accuracy of two classes with a threshold time of 18 months. The classification accuracy based on two classes increased approximately to 69%, as presented in Table 2, and as shown in Fig.6.

Second, the statistical and intensity texture features were extracted from each slice, and then the features of 155 slices for each sample were combined to obtain the feature vector. This texture features were used to train various MLs to produce the best prediction model. The accuracy of the classification based on three classification classes did not exceed 46% with 10-fold accuracy. The PCA method was applied to reduce the dimensionality of the features vector. However, the accuracy did not improve.

Third, the histogram features were extracted and used to train various ML for classifying the data, based on three classes. The best classification accuracy did not exceed 50%. Therefore, the data was classified into two classes, short-overall survivors and long-overall survivors, and these two classes were thresholded with a survival time of 18 months. The classification accuracy increased to 65% when using KNN and SVM classifiers, as shown in Fig.7. a and b. Also, to improve the classification accuracy, histogram and volumetric features were combined to produce a combined features vector. The combined vector was tested with classification based on two classes, as previously mentioned. However, the best accuracy for the combined features improved to only 68.4%, as shown in Fig.7.c.

Finally, the deep features were extracted from specific slices that clearly contained the three regions of HGG glioma, using pre-trained AlexNet. A total of 3,703 samples were prepared from the dataset as a training/validation dataset, and about 913 as a testing dataset. The accuracy with using 5 folds is very high, 91% for the Linear Discriminant classifier and 86.4% for the linear SVM classifier, as shown in Fig. 8. However, the accuracy of these models, using a separate test dataset, did not exceed 55%. Therefore, the best overall accuracy from the training/validation dataset, and the test dataset was 73% when using the Linear Discriminant classifier.

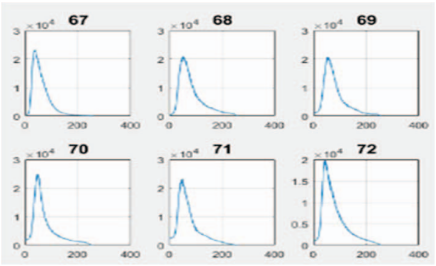

Fig. 4 Histogram of the brain region for samples 67 to 72.

T1	73%	17%	10%
T2	35%	24%	41%
T3	26%	21%	53%
	P1	P2	P3

(a)

T1	75%	8%	17%
T2	56%	24%	21%
T3	37%	13%	50%
	P1	P2	P3

(b)

Fig.5 Confusion matrix a) for nine features and an ensemble (RUS-boosted trees) classifier with 30-fold accuracy (50.8%) and b) for six features and a cosine KNN classifier with 12- fold accuracy (52.5%).

T1	79%	21%
T2	42%	58%
	P1	P2

79%	21%
58%	42%
TPr	FNr

Fig.6 Confusion matrix for nine features and logistic regression classifier with 5-fold accuracy (68.8%).

Table 1 Classification accuracy using volumetric and location feature based on three classification classes

# of Features	# of Folds	Best ML	Overall accuracy
9	5	Linear SVM	48.5%
9 with absolute to the location features	5	Coarse Gaussian SVM	49%
7 (*x* and *y* were removed)	5	Cosine KNN	50.9%
4 : the distance, R total tumor, R enhanced tumor, and R core	15% held out validation	Complex and median tree	62.5%

T1	71%	29%
T2	34%	66%
	P1	P2

T1	70%	30%
T2	40%	60%
	P1	P2

T1	80%	20%
T2	50%	50%
	P1	P2

Fig.7 Confusion matrix for hisogram features using: the left) weighted KNN, the middle) a median Gaussian SVM, and the right) combined features using linear SVM

T1	90%	4%	6%
T2	5%	92%	4%
T3	5%	4%	91%
	P1	P2	P3

(a)

T1	91%	6%	3%
T2	8%	83%	9%
T3	11%	6%	83%
	P1	P2	P3

(b)

T1	1441	66	92
T2	46	922	36
T3	50	44	1006
	P1	P2	P3

(c)

T1	1453	96	50
T2	82	835	87
T3	122	65	913
	P1	P2	P3

(d)

Fig.8 Deep-learning features of a confusion matrix, using a) TP FN rates for the Linear Discriminant , b) TP and false negative rates for linear SVM, c) observations for the Linear Discriminant, and d) observations for SVM.

Table 2 Classification accuracy using volumetric and location features based on two classification classes

# of Features	# of folds	Best ML	Overall accuracy
9-absolute to the location features	5	Logistic regression	68.8%
9-absolute to the location features	5	Quadratic Discriminant	68.8%

IV. Conclusions and Future Works

To improve the prediction accuracy regarding the overall survival for glioma patients in MRI, various types of the features were extracted and trained using various ML techniques. The best classification accuracy was achieved by using deep feature extraction based on pre-trained AlexNet and trained by Linear Discriminant. Although the texture features have been used widely in classification of many types of tumors, its accuracy did not exceed 46%. Histogram features achieved an accuracy of 68.5% for two classes of classification; this low accuracy could be due to noise. The model of the noise in the MRI images depended on the number of the coils in the imaging system. The noise is modeled as a Rician distribution in a single-coil system, and as a non-central chi distribution in a multi-coil (parallel) imaging system. From Fig.4, it was clear that the data that was provided was noisy data. The single-coil system assumes that the real part and imaginary part of an MRI image is an uncorrelated Gaussian distribution with a zero mean and equal variance. In fact, the Rician distribution is a special case of non-central chi distribution [20-22]. Therefore, one recommendation for future research would be to denoise the data and test the histogram features as well as other types of proposed features.

Furthermore, the texture features were extracted with specific slices that contained the three tumor regions for each sample. To improve the classification accuracy based on deep features, the specific patch size that contains the tumor could be used only to extract the 2D and 3D deep features, based on classification for three classes; then classification based on two classes also could be tested.

References

[2] American Brain Tumor Association (ABTA), "Brain tumor statistics," accessed Aug. 12, 2017. http://www.abta.org/about-us/news/brain-tumor-statistics/

[3] E. C. Holland, "Progenitor cells and glioma formation," J. Current. Opinion in Neurology, vol. 14, no. 6, pp. 683–688, 2001.

[4] H. Ohgaki and P. Kleihues, "Population-based studies on incidence, survival rates, and genetic alterations inastrocytic and oligodendroglia gliomas," J. of Neuropathology & Experimental Neurology, vol. 64, no. 6, pp. 479–489, June 2005.

[5] BH Menze *et al.,* "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", J. IEEE Transactions on Medical Imaging, vol. 34, no. 10, 2015, pp. 1993-2024, doi: 10.1109/TMI.2014.2377694 .

[6] S. Bakas *et al.*, "Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", J. Nature Scientific Data, Sept. 5 2017, doi:10.1038/sdata.2017.117.

[7] S. J. S. Gardezi, and I. Faye, "Fusion of Completed Local Binary Pattern Features with Curvelet Features for Mammogram Classification," Int. J. Applied Mathematics & Information Sciences, Appl. Math. Inf. Sci., vol. 9, no. 6, pp. 3037-3048 (2015).

[8] S. Pramanik, D. Bhattacharjee, and M. Nasipuri, "Wavelet based thermogram analysis for breast cancer detection", *Int. Symp. on Advanced Computing and Communication (ISACC) 2015*, Silchar, India, pp. 205-212, DOI: 10.1109/ISACC.2015.7377343.

[9] J. Sachdeva *et al.*, "Segmentation, Feature Extraction, and Multiclass Brain Tumor Classification." J. of Digital Imaging, vol. 26, no. 6 (2013), pp. 1141–1150, doi: 10.1007/s10278-013-9600-0.

[10] V.P. Gladis Pushpa Rathi, and S. Palani, "Brain Tumor Detection and Classification Using Deep Learning Classifier on MRI Images," Research Journal of Applied Sciences, Engineering and Technology, vol. 10, no. 2, doi: 177-187, May 2015.

[11] A. Faro *et al.*, "Statistical Texture Analysis of MRI Images to Classify Patients Affected by Multiple Sclerosis," *XII Mediterranean Conf. on Medical and Biological Engineering and Computing 2010*, pp. 272-275, May 27-30, 2010, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-642-13039-7_68.

[12] R. Paul *et al.,* "Combining Deep Neural Network and Traditional Image Features to Improve Survival Prediction Accuracy for Lung Cancer Patients from Diagnostic CT," *2016 IEEE Int. Conf. on Systems, Man, and Cybernetics SMC 2016*, October 9-12 2016, Budapest Hungary, pp. 2570- 2575, doi: 10.1109/SMC.2016.7844626.

[13] R. Paul *et al.,* "Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival among Patients with Lung Adenocarcinoma," Tomography : a journal for imaging research, vol.2, no. 4, 2016, pp.388–395, doi: 10.18383/j.tom.2016.00211.

[14] S. Liu *et al.*, "Prostate Cancer Diagnosis using Deep Learning with 3D Multiparametric MRI," *on Conf. Proc. SPIE Medical Imaging March 2017 vol. 10134, Medical Imaging 2017: Computer-Aided Diagnosis*, Orlando, Florida, United States, doi: 10.1117/12.2277121.

[15] D. Nie *et al.*, "3D Deep Learning for Multi-modal Imaging-Guided Survival Time Prediction of Brain Tumor Patients," *on Proc. 19th Int. Conf. Medical Image Computing and Computer-Assisted Intervention 2016*, MICCAI 2016, Athens, Greece, October 17-21, Part II, pp.212-220, Springer International Publishing.

[16] L. Zhao and K. Jia, "Multiscale CNNs for Brain Tumor Segmentation and Diagnosis," J. Computational and Mathematical Methods in Medicine, vol. 2016 (2016), 7 pages, Article ID 8356294, doi:10.1155/2016/8356294.

[17] S. Pereira *et al.*, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," J. IEEE Transection on Medical Images, vol. 35, no. 5, pp. 1240-1251, May 2016.

[18] N. Kumar *et al.*, "Convolutional Neural Networks for Prostate Cancer Recurrence Prediction," *on Proc. SPIE 10140, Medical Imaging 2017: Digital Pathology*, Vol. 10140, Orlando, Florida, United States doi: 10.1117/12.2255774

[19] S. Hu *et al.,* "Texture feature extraction based on wavelet transform and gray-level co-occurrence matrices applied to osteosarcoma diagnosis," *Proc. of the 2nd Int. Conf. on Biomedical Engineering and Biotechnology*, 11–13 October 2013, Wuhan, China , published in J. Bio-Medical Materials and Engineering vol. 24, no. 1, pp.129-43. doi: 10.3233/BME-130793.

[20] A. Krizhevsky *et al*., "ImageNet Classification with Deep Convolutional Neural Network*," on Proc. 25th Int. Conf. on Neural Information Processing Systems*, NIPS'12, pp.1097-1105, 2012.

[21] D. Drumheller, "General expressions for Rician density and distribution functions," J. IEEE Trans. Aerosp. Electron. Syst., vol. 29, no. 2, pp. 580–588, Apr. 1993.

[22] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, , vol. 34, no. 6, pp. 910–914, Aug. 1995.

[23] C. D. Constantinides, E. Atalar, and E. McVeigh, "Signal-to-noise measurements in magnitude images from NMR phased arrays," official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine. 1997, vol. 38, no. 5, pp. 852-857.