

Predicting Survival of Brain Tumor Patients using Deep Learning

Sharmila Agnal A, Arun Deepak C, Venkatesh J, Sudarshan S, Pranav A

Abstract: Deep Learning and advents in the field of machine learning, in general, has given rise to powerful classification algorithms applied to various real-world applications, hitherto requiring human experts. Consequently, the proposed system aims to automatically predict the survival of patients suffering from Glioma, a type of highly fatal brain tumor characterized by survival rates lower than two years. Until now, once a patient is diagnosed with Glioma it is the physician who provides the estimated number of days the subject would survive. However, the proposed system aims to automate the very process in order to obtain an unbiased prediction, bereft of any human error. Based on data consisting of Magnetic resonance imaging (MRI) images (four structural modalities), ground truth segmentation labels marking the region of interest (ROI) and accompanying information such as age and resection status of each patient, the system is expected to perform a classification task- high risk (less than 10 months survival), moderate risk (between 10 to 15 months survival) and low risk (greater than 15 months). While MRI image acquisition process can be complex, BraTS 2018 dataset fulfilled this vital requirement of the proposed system. The task entails two essential modules- feature extraction and classification. VGG16, a pre-existing Convolutional Neural Network (CNN) model, is employed to extract essential features from MRI images. Finally, an Artificial Neural Network (ANN) classifies the survival of a patient into the categories mentioned above. Prediction accuracy is measured using 5-fold cross-validation and a test dataset. The system achieved 100 % accuracy on training set itself, 88.8% using 5-fold cross-validation and 52% on an unknown test set, thereby, averaging 80.3% accuracy. While Glioma is, unfortunately, an extremely fatal condition, an accurate survival estimate can be a boon to the subject so as to fulfill one's commitments and bid adieu to their loved ones. Thus, unbiased opinion from a well-trained prediction model can be greatly applied in medical institutions to give their patients the correct estimate of their survival.

Index Terms: Deep Learning, Machine learning, Glioma, MRI, VGG16, CNN, ANN, 5-fold cross-validation.

I. INTRODUCTION

Advancements in hardware technology have opened up computational methods to solve complex biological problems

Revised Manuscript Received on April 07, 2019.

Sharmila Agnal A, Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.

Arun Deepak C, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.

Venkatesh J, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.

Sudarshan S, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.

Pranav A, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.

[1]. Novel techniques to solve medical imaging problems, medical image segmentation, in particular, have burgeoned due to many image segmentation challenges conducted to encourage automatic segmentation techniques. Multimodal Brain Tumor Image Segmentation (BraTS), Ischemic Stroke Lesion Segmentation (ISLES), and MR Brain Image Segmentation (MRBrainS) are just some of the ubiquitous contests that exist [2]. MRI has been attracting increasing attention in recent years as it is non-invasive and provides good soft tissue contrast [2]. BraTS challenge consists of two tasks based on a dataset consisting of magnetic resonance scanning (MRI) images in four different structural modalities, ground truth segmentation labels and survival information acquired from various institutions [3-4]. The dataset consists of patients suffering from Glioma, a type of tumor considered to be the most aggressive with overall survival not exceeding two years [5]. The first task demands the participants to segment the Glioma from MRI scan images based on the multimodal training data. The second task involves the prediction of overall survival of the patients. The proposed system focuses on the second task in detail. The dataset distributed as a part of BraTS 2018 contest is the latest one and is used in the proposed system.

II. EXISTING SYSTEM

The most essential tasks of the proposed system entail feature extraction and application of classification algorithms. Machine Learning and deep learning techniques have been widely used for such tasks in the existing systems. Lina Chato and Shahram Latifi [5] use BraTS 2017 dataset to predict overall survival of brain tumor patients. For the feature extraction task volumetric, statistical, histograms and deep feature techniques were employed. The classification algorithms such as SVM, k-nearest neighbors, linear discriminant tree were used for the prediction task. Lina Chato and Shahram Latifi [5] established that deep feature extraction using pre-trained Convolutional Neural Networks (CNN) was the most accurate with high efficiency. Linear Discriminant Classifier was found to be the best classifier. The accuracy using Alexnet for feature extraction and linear discriminant classifier using 5-fold validation was 91 % but did not exceed 55% with distinct test data. The average performance of their best feature extraction and classification algorithms combined was 73%. Deep feature extraction has two major advantages, firstly, deep learning can automatically find features from a given dataset for each specific application,



secondly, while traditional feature extraction is limited to prior knowledge, deep learning can find new features that are suitable to specific applications, but have never been previously discovered by researchers, explains Jianxin Wang and et al. [2]. The advantages are evident in [6] where Xinggong Wang and et al., in order to classify prostate cancer patients from patients with benign conditions, use both non-deep learning methods like SIFT image feature and deep learning methods. The AUCs were 0.84 for deep learning method and 0.70 for non-deep learning method, respectively, thus indicating that deep learning techniques gave much better results than non-deep learning methods. It is also observed that the above can be applied to various other image modalities like CT, MRI and PET scans of other organs. Transfer Learning involves the use of the output of one machine learning algorithm as input to another one. Guan Wang, Yu Sun, and Jianxin Wang [7] inspired by the deep learning breakthrough in image-based plant disease recognition, their work proposed deep learning models for image-based automatic diagnosis of plant disease severity. They found the performance of fine-tuned models was superior to that of models trained from scratch. The best result was achieved by the VGG16 model, with an accuracy of 90.4%. The results indicated that transfer learning alleviates the problem of insufficient training data. For comparison, an ANN model was trained by SGD optimizer end-to-end on the training set. Test accuracy of 31% was achieved, which was basically random guessing. Without the convolutional feature extractor, the ANN could not extract local correlations and learn discriminative features from the images. Rahul Paul and et al. [8] also obtain high accuracy in predicting lung cancer using VGG16.

Ali Ari and et al [9] proposed a local receptive field extreme learning machine which takes a single hidden layered feedforward neural network input weights randomly and output weight analytically for classification of type of tumor whether it is benign or malignant with preprocessing for reducing noise in a given MRI image and watershed segmentations which checks for the topology of the cranial MRI is used to detect the tumor in a given MRI image. This method of classification had an efficiency of 95% compared to other algorithms such as statistical features (93%) and Gabor-wavelet features (90%). Dong Nie and et al [10] proposed a deep learning framework to automatically extract features from MRI images of high-grade Glioma patients. They use three-dimensional Convolution Neural Network (CNN) as well as create an architecture for multi-channel data. The CNN extracts its own features from the MRI images and doesn't require to be trained from the training dataset. A support vector machine (SVM) is also created and trained to predict the survivability of patients with high-grade gliomas. SVM analyzes data used for classification and regression analysis. With this deep learning framework, they achieved maximum accuracy of 89.9% throughout the training sets.

Fabian Isensee and et al [11] proposed a Segmentation Algorithm in the form of Convolutional Neural Network (CNN). The Algorithm is modified to maximize the

performance for segmentation of brain tumor. In order to reduce class imbalance, they use dice loss function and augment the data to prevent overfitting. They also train random decision forest and multilayer perceptron on extracted features describing tumor subregions to predict the survivability of brain tumor patients. They achieved a maximum of 52.6% accuracy in predicting the survivability and also scored 0.496 in Spearman correlation coefficient and a mean error of 209697. The Brain tumor Segmentation using the CNN Algorithm helped in achieving a Dice score of 0.858 for the whole segment, 0.775 for the core part of the segment and 0.647 for the segment of the enhanced tumor.

Harry B Burke and et al [12] proposed the use of Artificial Neural Network (ANN) to predict survivability of patients with Breast and Colorectal Carcinoma. Artificial Neural Network consisting of three layers for accurate predictions. The system also employed an Artificial Neural Network with backpropagation training, the maximum likelihood estimation function, and a gradient descent algorithm. Breast and colorectal carcinoma patient care evaluation and the National Cancer Institute's Surveillance, Epidemiology and End Results data set were used for training the Algorithm and predicting the output. A comparative study between the TNM staging system and ANN was done to check which one has better accuracy in predictability. The TNM staging system throughout the training shows an accuracy of 47% whereas placing the dataset in ANN produced an accuracy of 74%. Shobana and et al. [13] focus on diagnosing a tumor using 3 stages. Firstly, the image is sharpened, and the noise is removed. In the second phase, DCT and DWT are used for feature extractions. And finally, PNN is used as a classifier. The drawback is that DCT is less efficient compared to VGG 16 and ANN is a better classifier than PNN. Vaibhavi Solanki and et al. [14]. discuss the hybrid approach of two different classification algorithms like SVM-RBF to yield higher accuracy values than SVM. In the initial stage of the system, the feature extraction is done using GLCM and the features are inputted to the SVM classifier. The study also discusses that SVM and KNN hybrid approach can give even better results than that of the SVM-RBF approach.

Wenbin Zhang and et al. [15] proposed an algorithm using the KNN on clinical parameters such as age and severity of the tumor. The feature extraction was accomplished using the technique known as fast correlation based filtered approach. Zhang listed all the features from the FCBF approach in the descending order of the symmetrical uncertainty values. Some features with lower threshold values were eliminated. Hussam Qassim and et al. [16] discuss how a compressed form of VGG16 model is smaller in size and has a faster training time. It also uses residual learning for improving generalization and convergence and solving the problems of degradation. The study showed that the above model is 88.4% smaller and 23% faster than the actual VGG model. Dr. Basavaraj Amarapur [17] suggested the use of the Artificial Neural Network as a classifier to yield better results than that of other



hybrid models. The proposed system yielded around 85% accuracy. The performance was examined with 40 trained images and 60 test MRI images. The accuracy was calculated relative to the confusion matrix.

III. PROPOSED SYSTEM

A. System Overview

The first step of the system focuses on collecting the dataset with the required features and further refine it through preprocessing to obtain a clean input for machine learning. Since 3 dimensional MRI images form a vital component of the dataset, there arises a need for a tool to manipulate MRI images. The proposed system uses the python based nibabel library to access the MRI images obtained from BraTS2018 [3] in NifTI (.nii) format. Then using matplotlib library of python the data is visualized. The essential steps, however, include feature extraction and designing the classification model. As observed in [2], [5-8], deep feature extraction mechanisms using a pre-trained CNN model has been more efficient to identify features of a given image than traditional methods as the task is automatic and it is able to pick up on even application specific features never previously discovered by researchers [2]. In [7-8] best results were observed through transfer learning. Moreover, the dataset includes both image and non-image data making transfer learning important for the proposed system. The proposed system uses a pre-trained CNN model, VGG16[7-8] for feature extraction, which along with other features like age is fed as input to an Artificial Neural Network classifier through transfer learning. Each step is described in depth in the following sections.

B. Dataset

The system uses data from the BraTS 2018. The file format of MRI images in the dataset is NIFTI with extension nii. The MRI images are represented in four different modalities- T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes, native (T1), T2-weighted (T2), and post-contrast T1-weighted (T1Gd). The dataset was obtained from various institutions with different clinical protocols and numerous scanners. With the approval of expert neuro-radiologist and using standard annotation protocols, the data set was segmented by 1-4 raters manually. The necrotic and non-enhancing tumor core (NCR/NET — label 1), the peritumoral edema (ED — label 2) and the GD-enhancing tumor (ET — label 4) include the different annotations used. The training data included, for each subject, the 4 structural modalities, ground truth segmentation labels and accompanying survival information, age, and resection status.

In order to obtain the dataset [4] following steps are to be followed:

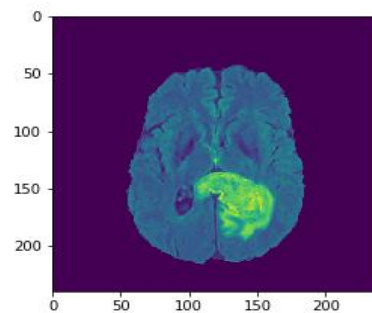
- Create an account in CBICA's Image Processing Portal (ipp.cbica.upenn.edu) and wait for its approval.
- Once the IPP account is approved, login to ipp.cbica.upenn.edu and then click on the application "BraTS'18: Data Request", under the "MICCAI BraTS 2018" group.
- Once the request is recorded, you will receive an email pointing to the "results" of your submitted job. You need to

login to IPP, access the "Results.zip" file, in which you will find the file "REGISTRATION_STATUS.txt" that will provide the links to download the BraTS 2018 data.

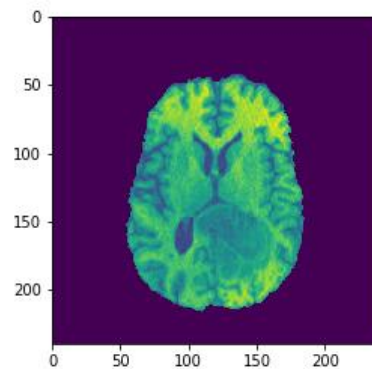
The dataset includes 163 records of Glioma patients with a unique ID. Each has MRI images in four different modalities, a ground truth segmentation label and a separate CSV file consisting of the features- age, overall survival and resection status- with the unique ID being the association. 121 records were separated as training data and the rest for testing.

C. Data Visualization

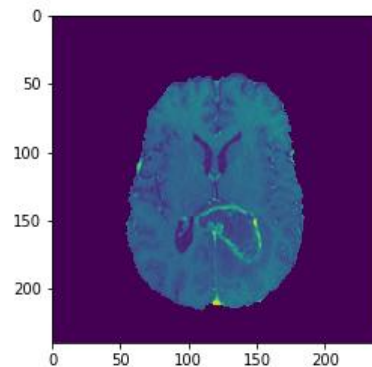
The nibabel library of python is instrumental in accessing Nifti format files. The files loaded using nibabel are then plotted using the matplotlib library of python. The result obtained by implementing the above library on the dataset is illustrated in Fig.1 and Table 1.



(a)



(b)



(c)

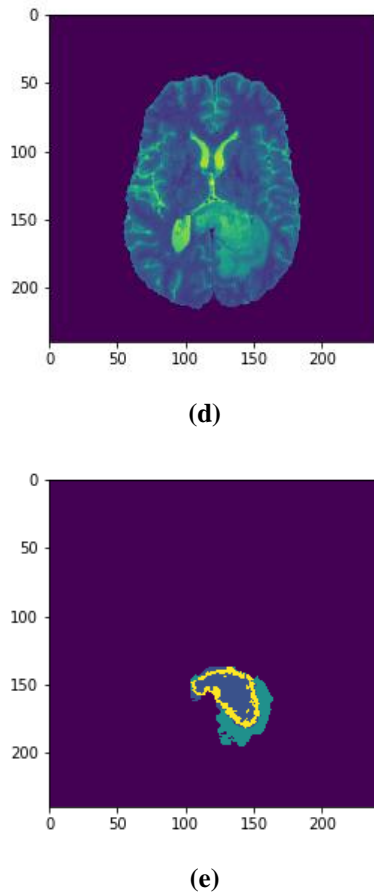


Fig. 1 Visualization of MRI images from BraTS 2018 of the subject with ID Brats18_CBICA_AWG_1 describing: a) T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes, b) native (T1), c) post-contrast T1-weighted (T1Gd), d) T2-weighted (T2), e) ground truth segmentation labels.

Table 1 Record of subject Brats18_CBICA_AWG_1 in BraTS 2018.

BraTS18ID	Age	Survival	Resection Status
Brats18_CBICA_AWG_1	55.532	180	GTR

D. Feature Extraction

Feature Extraction is the process of reducing an initial set of raw data to a smaller, understandable and manageable groups known as features or feature vectors for processing while accurately describing the original data set. Feature extraction uses various techniques and methods to construct numerous combinations of a given data to get around the problem of the sluggishness of the system to understand the data, as a whole, while still describing the original data with great accuracy. The most notable application of feature extraction is in Image Processing. Features such as Pixel Density, Edge Detection, Shape-based Detection, Template Matching, Motion Detection, etc. are extracted from an

image to analyze, process and predict various other parameters that may be required by the various Classification and Neural Network Algorithms.

VGG, also known as The Visual Geometry Group, is a Convolutional Neural Network (CNN) model for image recognition. It is used for feature extraction of images too. VGG 16 is the model used in the proposed system for feature extraction. It is a VGG model with 16 weighted layers. VGG 16 takes an input image of size 224*224*3. VGG network is characterized for its simplicity. It uses only 3x3 convolutional layers stacked on top of each other in increasing depths. It reduces the volume of the image by max pooling. It can classify image up to 1000 object categories. VGG 16 is available in Keras deep learning library for feature extraction and classification. It uses pre-trained weights for feature extraction and classification to train the model and once it is trained it can be used for newer weights for prediction and classification. The model only takes an image of size 224*224, however, an image can be easily resized to 224*224 using functions in Keras library. Once resized it does certain pre-processing of the image as the network may expect one or more images as input and adds the desired dimensions. After the pre-processing, it makes the predictions of the probability of the image belonging to each of the 1000 known objects types. Once the prediction is made it can return a list of classes and their probabilities of top 3 objects that may be present in the image. The output of the whole process is the identified image with its prediction accuracy. VGG 16 can give an accuracy of up to 95%. VGG 16 architecture is illustrated in Fig.2 [18]. For feature extraction, however, the top two layers are excluded. The output from the layer with 7*7*512 values is then fed into a classification algorithm using transfer learning.

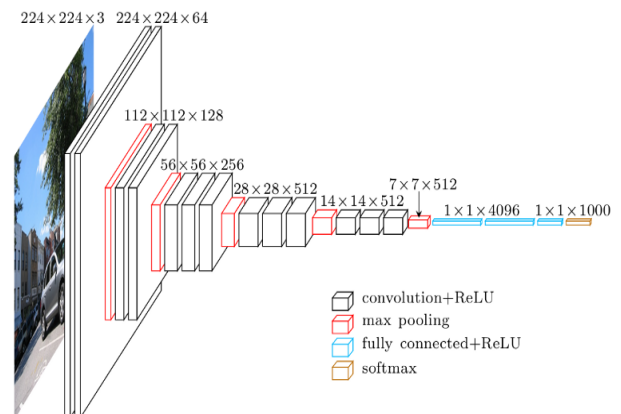


Fig. 2 VGG16 architecture

Transfer learning is a field in machine learning that stores knowledge obtained while solving a certain problem and applying that knowledge to a different but related problem. This method helps in training a Deep Neural Network with comparatively fewer data. An example would be using the knowledge gained while learning to recognize cars being applied when recognizing trucks. Implementation of transfer learning entails usage of pre-existing models already trained on image or



language datasets thereby completely bypassing the need to develop a neural network from scratch for each related application.

Two common Transfer Learning approaches are as follows:
Develop Model Approach

This approach involves selecting a problem which has relations or similarities with the current problem, developing a source model for solving it and then reusing that model as the starting point to solve your main task.

Pre-trained Model Approach

This approach involves selecting a pre-existing model from a group of candidate models to be reused as the starting point for your task of interest.

The proposed system uses VGG16, for extracting features and that in turn is used as a starting point for the classification task. The existing systems reflect a strong inclination towards transfer learning giving high accuracy and hence the proposed system uses the same.

E. ANN Classifier

The proposed system uses an artificial neural network to classify the overall survival of the subjects as high risk (<10 months), moderate risk (>10 months and <15 months) or low risk (>15 months). Transfer learning is used to combine the useful features extracted using VGG16 with the age of the patient as the input to the classifier. The labels represent survival in terms of values 0,1,2, where 0 represents high risk, 1 moderate risk, and 2 low risk.

The output from VGG16 consists of 25088 features from MRI images (7*7*512). Further cleaning and feature selection enable a reduction to 4427 features. Reducing the number of features by removing useless ones can greatly improve the performance of ANN classifier. Then, the patient age feature is appended to the existing features to form a dataset with 4428 features. The model has an input layer with 4428 inputs, two hidden layers with 16 activation units each, with rectified linear unit (relu) activation. Finally, the output layer has 3 activation units representing the three outcomes. The output layer uses a SoftMax activation in order to obtain the output as probabilities for classifying into one of the categories. The loss evaluation is done using categorical cross-entropy as this is a classification problem. Gradient descent is used as the optimizer while calibrating weights and bias. Fig. 3 summarizes the ANN model.

Layer (type)	Output Shape	Param #
dense_97 (Dense)	(None, 16)	70864
dense_98 (Dense)	(None, 16)	272
dense_99 (Dense)	(None, 3)	51
Total params: 71,187		
Trainable params: 71,187		
Non-trainable params: 0		

Fig.3 Artificial Neural Network Classifier

F. Proposed Algorithm

Fig. 4 depicts the proposed algorithm. The six major steps are as follows:

Step-1: Data Collection

Input parameter: Username and Password.

Process: Firstly, an account is created in CBICA's Image Processing Portal. Once the account is created, the data is requested using the MICCAI BraTS'2018 tab. Once the data is requested, the UPENN takes a couple of days to approve the account. Once the account is successfully approved, the UPENN sends the training data in the form of a zip file. The user can download the file through the provided link.

Output: The download that is provided after the successful approval of the account.

Step-2: Data Visualization

Input Parameter: The data set is the input for this module.

Process: Firstly, the libraries are imported into the program that is accessing the data set. The nibabel library is used for opening the image files in NIFTI (.nii) format. The image can be viewed through the program. And the matplotlib is used for making graphs from a given raw data. The raw data is then isolated from the data set and stored in a csv file.

Output: MRI images in 4 different modalities and CSV file.

Step-3: Feature Extraction

Input: VGG model and Keras library.

Process: Firstly, the keras library is imported into the training data set. Along with the library, the VGG16 model is also imported. The VGG model is fed with the training data to extract the features. The VGG16 uses 3x3 convolutions layers and 2x2 max-pooling layers and are actively connected at the end. It takes the input image of size 224*224*3. The image is RGB. Once the features are extracted from the network, they are added to the list. The program now has 25088 features of the image.

Output: 25088 unique features.

Step-4: Artificial Neural Network Classification

Input: Extracted 4428 features after feature reduction

Process: The extracted features are inputted to the ANN model. The ANN model has four layers; input layer, two hidden layers (relu activation) and output layers (3 units softmax activation). The 30 epochs are enabled with early stopping. The model summary is generated. The generated model is compiled with desired loss and optimization metrics. The training data is fitted, and the model is saved in the destination folder. The ANN parameters like weight and bias are modified to improve accuracy based on performance. The ANN predicts the risk in three categories; high risk, moderate risk, and low risk.

Output: Risk category is predicted.

Step-5: Performance Analysis

Input: Training set and Test set

Process: The training set is considered and evaluated. The accuracy on the training set is noted. The training set is then evaluated using 5-fold cross-validation and the accuracy is noted. The test set is then evaluated, and the accuracy is noted. The confusion matrix is generated for the above tests.

Output: Confusion Matrix based on all the tests.

Step-6: Tkinter GUI Application

Input: Patient ID

Process: The tkinter app is opened and the user is allowed to select the patient id from the drop-down box. The patient id is then taken by the application and the prediction takes place. Once the process is over, the program outputs its



prediction in a new window. The GUI application also allows us to visualize the patient data. The MRI images can be opened through the application and visualized manually. The application allows the user to check the performance of the particular iteration. The accuracy achieved by the algorithm is displayed and the confusion matrix is displayed for the very evaluation of the data set.

Output: The tkinter app with prediction and visualization features.

IV. EXPERIMENTAL RESULTS AND EVALUATION PERFORMANCE

The training data was fit into the ANN model designed, to calibrate the weights and bias.

each case, as observed in Fig.5. While there are 53 high risk cases present in the training set, there are only 31 moderate risk cases and 37 low risk cases. It is observed that a massive boost in the accuracy, to the tune of 80% is obtained as against 30% simply because that particular case has 20 data records more in the training set. Thus, the overall performance on unknown test set can be improved by increasing the representation of each case in the training set. Combining the 88.8% accuracy of 5-fold cross-validation, 52% accuracy on the test set and 100% on the training set, the average prediction accuracy of the system is 80.3% on just 121 samples of training data. Table 2 summarizes the system performance. It is clearly elucidated in the table what the different testing strategies are and how well the system performs in the respective category.

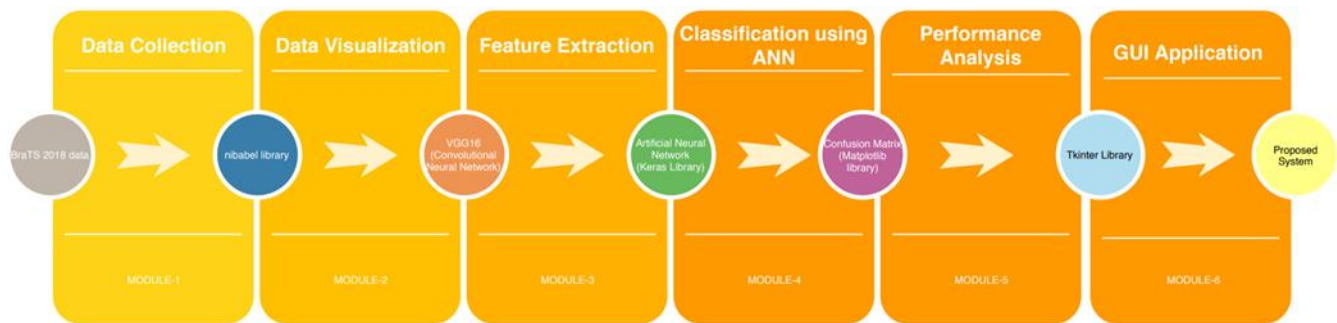


Fig. 4 Proposed Algorithm

While 30 epochs were passed, an early stopping mechanism was designed with a patience limit of 10 and comparison of up to four decimal places. After training when the training set itself was put to test a 100% accuracy was obtained as illustrated in Fig.6. Validation of the model was performed in two different ways- 5-fold cross-validation and a separate test set. Deep feature extraction, deep learning in general, works best with a large dataset, however, there is a dearth of data available to the system as the BraTS 2018 dataset is quite small. Lina Chato and Shahram Latifi[5] have prepared a custom dataset consisting of 3703 samples from the given 163 samples. However, when Lina Chato and Shahram Latifi[5] used Alexnet for feature extraction and linear SVM classifier, they obtained 5-fold cross-validation of 86.4%. When the proposed system used VGG16 for feature extraction and ANN classifier, even though the training set just had 121 samples, it obtained a 5-fold cross validation score of 88.8%. When dimensionality reduction using Linear Discriminant Analysis was applied, no improvement was observed.

On the other hand, when a separate test dataset was applied to the system, the accuracy did not exceed 52%. However, the confusion matrix in Fig.8 reveals that the system predicted high risk cases with high accuracy (about 80%). But the prediction accuracy of moderate and low risk cases is low (38% and 29% respectively). The disparity can be directly attributed to the number of training examples available for

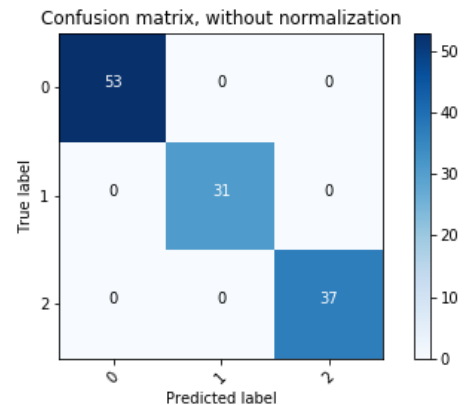


Fig. 5 Confusion matrix of predictions on the training set (Number of Instances).

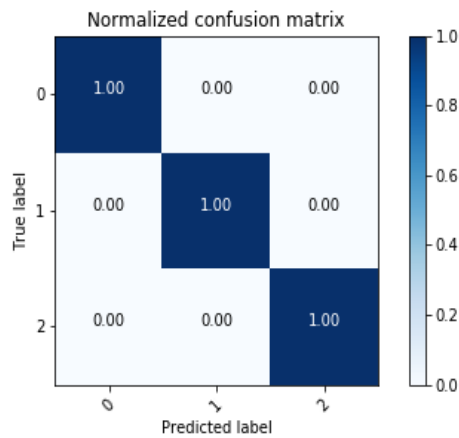


Fig. 6 Confusion matrix of predictions on the training set (Probability).

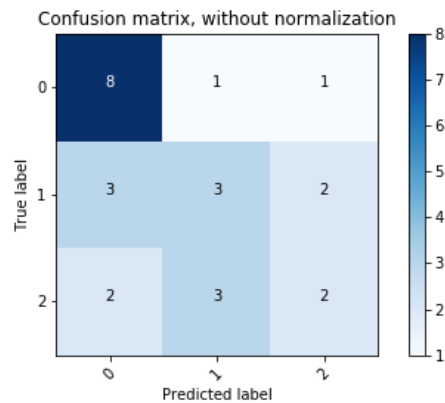


Fig. 7 Confusion matrix of predictions on the test set (Number of Instances).

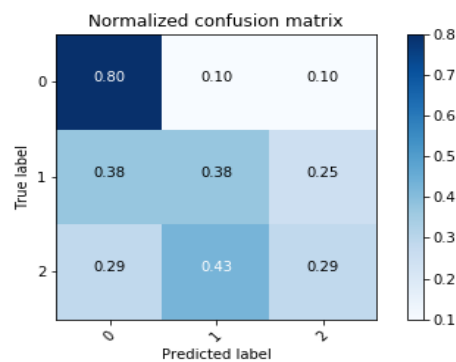


Fig. 8 Confusion matrix of predictions on the test set (Probability).

Table 2 Performance of System

Testing Strategy Used	Accuracy
Training data test	100%
5-Fold Cross Validation	88.8%
Unknown Test data	52%
Overall System	80.3%

Performance

V. PERFORMANCE ANALYSIS OF THE PROPOSED SYSTEM

Various feature extraction and classification algorithms have been employed by researchers in the existing systems. Key to a better solution lies in finding the perfect combination of the above algorithms. One thing commonly observed across various existing systems is the fact that Deep Learning feature extraction has shown the highest accuracy.

Fig.9 shows that when existing systems' performance are measured using 5-Fold Cross Validation technique the feature extraction using volumetric performs poorly with majority of Machine Learning algorithms. Those existing systems struggle to reach 50% accuracy. However, feature extraction using deep learning has shown better performance and when employed with Linear SVM Classifier gave an 86.4% accuracy. Fig.10 depicting overall system performance only reemphasizes the superiority of deep learning feature extraction mechanism. While Statistical & intensity features stands at 46 % accuracy, histogram features too barely reach 50%. However, features learnt through deep learning touches 73%.Fig.9 and Fig. 10 reveal that the proposed system, represented by a golden legend in Fig.9 and Fig. 10, has outperformed the existing system both in terms of 5-Fold Cross-Validation and overall as a system. The proposed system, using deep feature extraction (pretrained CNN called VGG16) and artificial neural network for classification, marked an 88.8% accuracy using 5-Fold Cross Validation and 80.3% as a system on the whole. The proposed system attained a 100% accuracy on training data itself, which is notable considering the fact that it is a model-based machine learning algorithm and there is no direct instance-based comparison. In fact, the system adjusts its weights and bias to generate a generalized model, implying that the model has trained itself well. The overall test set accuracy could only reach 52%, however, considering the fact that majority of algorithms in the existing system have attained accuracies less than or equal to 50 %, the proposed system has outperformed the existing even in this category.

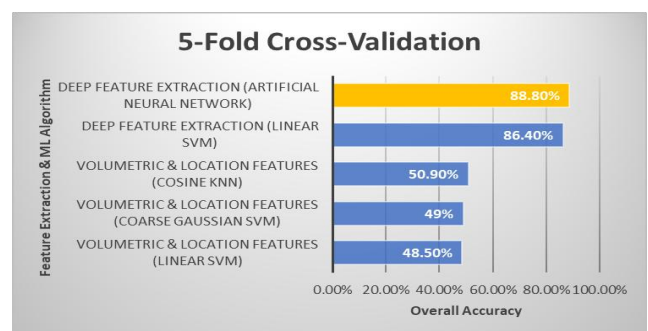


Fig. 9 Five-Fold Cross-Validation on Existing vs Proposed System.



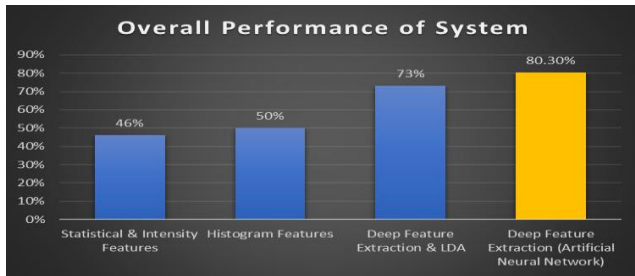


Fig. 10 Overall Performance of Existing vs Proposed System.

VI. CONCLUSION AND FUTURE WORK

It is evident that computational methods for solving biological problems, especially medical imaging, can be quick and cost-effective. As observed earlier, one of the major challenges in medical imaging is that the data is mostly images and therefore requires efficient feature extraction mechanisms. The existing systems pointed out that deep learning, CNN in particular, possesses elegant feature extraction mechanisms, including such features which researchers could not obtain through traditional methods. Further, it was also established that transfer learning provided high accuracy and therefore the proposed system imbibed the transfer learning concepts.

The paper began with a brief introduction of the problem at hand, in addition to giving an intuition on the BraTS 2018 challenge. Further, the existing systems to solve the particular problem were thoroughly analyzed by using 18 different references. What was universally observed, in all the references, was that best performance was obtained through deep learning-based feature extraction mechanism, which was then, after using transfer learning technique, used for classification. Then the proposed system was scrupulously explicated, including the various modules and algorithms required namely- Data Collection, Data Visualization, Feature Extraction, ANN Classification, Performance Analysis and GUI Application. After a step by step account of the implementation process there comes the result analysis.

To sum up the evaluation result, the study showed that even with a very small dataset of 121 training samples the system achieved 100% accuracy on training set itself, 88.8% using 5-fold cross-validation and 52% on an unknown test set, thereby, averaging 80.3% accuracy. From the confusion matrix it was inferred that while test set label prediction of high-risk patients was highly accurate touching 80%, due to the paucity of data representing moderate and low risk patients, the predictions were not very accurate for those cases.

A larger dataset can improve the performance of the system, especially to make the system learn about moderate and low risk patients, whose representation in the current system is low. Better feature extraction algorithms, data preprocessing are other factors to work upon for better results. Although, as established by the performance analysis section, the proposed system outshone the existing system both in terms of 5-fold cross validation and overall performance with an 88.8% and 80.3% respectively, there is still scope for improvement in terms increasing the accuracy in the future.

REFERENCES

1. Mahmud, M., Kaiser, M. S., Hussain, A., & Vassanelli, S. (2018). 'Applications of deep learning and reinforcement learning to biological data'. *IEEE transactions on neural networks and learning systems*, Vol.29, No.6, pp.2063-2079.
2. Liu, J., Pan, Y., Li, M., Chen, Z., Tang, L., Lu, C., & Wang, J. (2018). 'Applications of deep learning to MRI images: A survey'. *Big Data Mining and Analytics*, Vol.1, No.1, pp.1-18.
3. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... & Lanczi, L. (2015). 'The multimodal brain tumor image segmentation benchmark (BRATS)'. *IEEE transactions on medical imaging*, Vol.34, No.10, pp.1993-2024.
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., ... & Davatzikos, C. (2017). 'Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features'. *Scientific data*, Vol.4, pp.170117.
5. Chato, L., & Latifi, S. (2017). 'Machine learning and deep learning techniques to predict overall survival of brain tumor patients using MRI images'. In *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)* (pp. 9-14). IEEE.
6. Wang, X., Yang, W., Weinreb, J., Han, J., Li, Q., Kong, X., ... & Wang, L. (2017). 'Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning'. *Scientific reports*, Vol.7, No.1, pp.15415.
7. Wang, G., Sun, Y., & Wang, J. (2017). 'Automatic image-based plant disease severity estimation using deep learning'. *Computational intelligence and neuroscience*, 2017.
8. Paul, R., Hawkins, S. H., Balagurunathan, Y., Schabath, M. B., Gillies, R. J., Hall, L. O., & Goldgof, D. B. (2016). 'Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma'. *Tomography*, Vol.2, No.4, pp.388.
9. Ari, A., & Hanbay, D. (2018). 'Deep learning based brain tumor classification and detection system'. *Turkish Journal of Electrical Engineering & Computer Sciences*, Vol.26, No.5, pp.2275-2286.
10. Nie, D., Zhang, H., Adeli, E., Liu, L., & Shen, D. (2016). '3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients'. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 212-220). Springer, Cham.
11. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., & Maier-Hein, K. H. (2018). 'Brain tumor segmentation and radiomics survival prediction: contribution to the BRATS 2017 challenge'. *CoRR abs/1802.10508*.
12. Burke, H. B., Goodman, P. H., Rosen, D. B., Henson, D. E., Weinstein, J. N., Harrell Jr, F. E., ... & Bostwick, D. G. (1997). 'Artificial neural networks improve the accuracy of cancer survival prediction'. *Cancer*, Vol.79, No.4, pp.857-862.
13. Shobana, G., & Balakrishnan, R. (2015). 'Brain tumor diagnosis from MRI feature analysis-A comparative study'. In *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIECS)* (pp. 1-4). IEEE.
14. Solanki, V., Patel, M. V., & Pati, M. S. (2018). 'Brain MRI Image Classification using Image Mining Algorithms'. In *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 516-519). IEEE.
15. Zhang, W., Tang, J., & Wang, N. (2016). 'Using the machine learning approach to predict patient survival from high-dimensional survival data'. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1234-1238). IEEE.
16. Qassim, H., Verma, A., & Feinzimer, D. (2018). 'Compressed residual-VGG16 CNN model for big data places image recognition'. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 169-175). IEEE.
17. Amarapur, B. (2017). 'An Automated Approach for Brain Tumor Identification using ANN Classifier'. In *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)* (pp. 1011-1016). IEEE.
18. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). 'A review of the use of convolutional neural networks in agriculture'. *The Journal of Agricultural Science*, Vol.156, No.3, pp.312-322.