International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

Classification of Tumors from Brain MR Slices using PCA for Discriminative Models

Parvathy Jyothi
School of Computing
Kalasalingam Academy of Research and Education
Krishnankoil,India

Robert A Singh
School of Computing
Kalasalingam Academy of Research and Education
Krishnankoil,India

***Abstract*— Magnetic Resonance Imaging is a widely used non-invasive tool for tracking down abnormalities in human body. It is necessary to detect any deviation in shape or volume of anatomical structures as early as possible to reduce death rate. MRI provide medical images with superior resolution especially for soft tissues. Manual segmentation and classification of tumors is a time-consuming task and the effectiveness depends on the radiologist's efficiency. Machine learning plays a vital role in automatic classification of brain tumors. In this paper, a comparison study is done on machine learning models namely Support Vector Machine and Random Forest Classifier to grade brain tumors in MR images. The data is collected from publicly available dataset for conducting experiments. Principal Component Analysis (PCA) is used to extract features from the input brain MR images. The machine learning models classify brain MR slices into three categories namely Meningioma tumor, Glioma tumor and Pituitary tumor. The proposed system records classification accuracy of 90%. Other performance metrics used in the study are precision, recall, F1-score and confusion matrix.**

Keywords*— *Machine Learning, Principal Component Analysis, Support Vector Machine, Dimension Reduction, Glioma Tumor, Meningioma Tumor, Pituitary Tumor

I. INTRODUCTION

The various modalities used for learning brain tissues and activities include Computed Tomography (CT), Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), Magnetic Resonance Imaging (MRI) etc. MRI works on the principle of radio waves and disciplined magnetic field. When the radio waves are switched off, different brain tissues rest at different rates and are displayed as T1 weighted MRI, T2 weighted MRI, FLAIR (Fluid Attenuated inversion recovery) and contrast enhanced CE-T1W MRI.

U.S news reports says that the growth rate of brain tumor is increasing 12% every year [1]. In general, the two categories of tumors are benign tumor and malignant tumor where the former is less harmful and the latter is deadly. World Health Organization recorded brain tumors into 120 classes of tumors. They are ranked from low risk (grade I) to high risk (grade IV). The prevailing tumors are Glioma, Meningioma, Pituitary and Astrocytoma [2]. Glioma tumors represent about 52 percent of all dominant tumors whereas Pituitary tumors make 15 percent. Hence early detection of brain tumors are pivotal for prognosis.

For huge volumes of MRI data, manual classification is time consuming. Machine learning approaches contribute much to the field of medical image analysis [3], [4], [5] during clinical prognosis. Traditional machine learning techniques includes multiple steps like image preprocessing, extracting features from input data, reducing the dimension of data, segmenting region of interest, classifying data and many more. The crux is feature extraction as the accuracy of classification task highly relies on it. The significant features of brain tumors are correlated to tumor location, shape, size and texture [6].

The layout of this paper is organized as follows: Section 2 focuses on the existing research and study made in the field of brain tumor classification. A detailed description about the proposed work is given in section 3. Dataset is discussed in section 4. Section 5 discuss the evaluation metrics used in the current work. The experimental results and comparison are shown in section 6. In the final section, conclusion and future work is presented.

II. LITERATURE SURVEY

Machine learning models are getting popular these days through their efficiency and exactness in making predictions automatically. Generally, these models falls into two categories namely discriminative and generative models [11]. Generative models look into the prior knowledge and distribution of data. They are unsupervised learning algorithms that depends on Bayes theorem. Hidden Markov Model [12], Generative Adversarial Networks [13] are some examples of generative models.

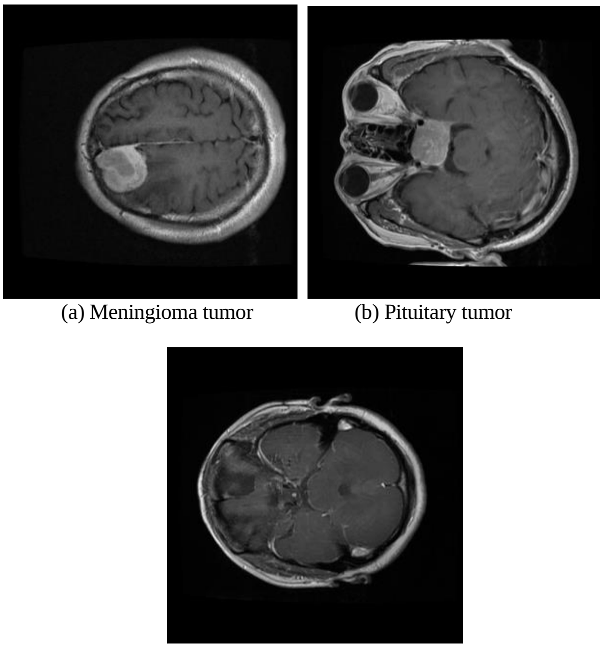
Discriminative models relies on hand crafted features of input data. In those models, features are used for training the network in an assumption that the feature vectors are capable of representing the input data very well. Optimization methods like Principal Component Analysis can be used to reduce the number of features for the productive usage of memory. Techniques like Support Vector Machine, Decision trees, Random Forest classifier, Fuzzy C-Means clustering belongs to these models.

Hebli Amruta et al. [14] developed a method to detect benign and malignant tumors from brain MR images. Initially, tumors are segmented using morphological operations. Discrete Wavelet Transform is used to extract features and PCA removes redundant information. The reduced feature data is finally fed to SVM for the classification task. Tumor classification using Gabor Wavelet Transform and SVM is presented in [15]. GWT is used to extract features and multi kernel SVM is employed for the classifying tumors.

Mudgal et al. [16] used Extreme Gradient Boosted decision trees for classifying tumors along with GLCM as feature extractor. The method detect tumor with the help of Marker-Controlled Watershed Transform. In another work [17] developed by Anitha R, image features like angular second moment (ASM), inverse difference moment (IDM), variance inertia, energy, dissimilarity, and homogeneity extracted using grey level co-occurrence matrix (GLCM) are fed to a Random Forest classifier. The classification result is binary which categorize input either to normal or abnormal.

III. PROPOSED WORK USING PCA

The initial goal of the proposed work is to classify brain MR images into Meningioma, Pituitary and Glioma tumors. Fig (1) (a), (b), (c) illustrates different MRI tumor slices from the dataset [10] used in this method. The flow chart of the proposed framework is shown as Fig (2). First, the input image is preprocessed so as to prepare it for classification. The preprocessed images are directed to a feature extractor. Here Principal Component Analysis (PCA) is employed for feature extraction and reduction. The extracted features are given to two classifiers namely Support Vector Machine (SVM) and Random Forest classifier. The experimental results shows that SVM gives effective classification of brain tumors.



(a) Meningioma tumor

(b) Pituitary tumor

(c) Glioma tumor

Fig.1 Different brain tumors in T1 weighted contrast enhanced MRI

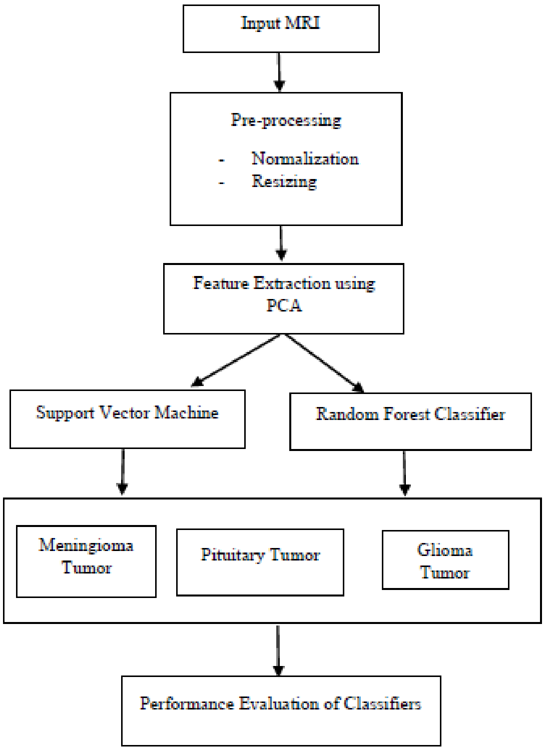


Fig.2 Flow chart of the proposed system

A. Pre-processing Input MR Image

Before supplying the images into the proposed classification network, a pre-processing is compulsory to follow input formats. The intensity values across MRI differ considerably when MR scans are collected from different scanners. Also, noise present in the captured MR scans may degrade fine details, blur edges etc. Hence it is necessary to normalize the input images. Intensity normalization is achieved by implementing min-max normalization that scale intensity value to [0, 1]. The min-max normalization is defined as (1). Here f'(x, y) represents the normalized intensity value, V_min and V_max are the minimum and maximum value in the image f.

$$f'(x,y) = \frac{f(x,y)-\text{V\_min}}{\text{V\_max}-\text{V\_min}} \quad (1)$$

After normalization, we resize the normalized image into 256x256. Resizing the MR images accelerates training process and reduce memory reservation. In fact, the size of MR images in the CE-MRI dataset [10] is 512x512 and the considered size for further classification is 256x256. A normalized resized image is shown in Fig (3).

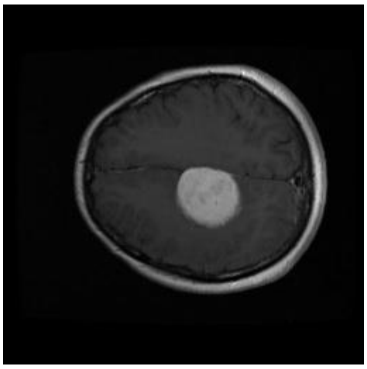


Fig.3 Normalized MR Image

B. *Feature Extraction Using Principal Component Analysis*

In machine learning, different methods exists for extracting features from the input data. Discrete Wavelet Transform (DWT) [7], Grey Level Co-occurrence Matrix (GLCM) [8] and Principal Component Analysis (PCA) [9] are some among them. In DWT, from the subclass of DWT coefficients, essential features from the tumor portion are extracted for further processing. In a gray scale image, the second order statistical information between neighboring pixels can be portrayed using GLCM. Using GLCM, features including contrast, correlation, energy, entropy, homogeneity and many more can be collected.

In the current work, PCA is used for feature extraction and dimension reduction. PCA reveals essential relationships among data, quantifies them and retain only principal components for classification task. The correlation among features are spotted using covariance matrix. The diagonal elements of covariance matrix gives the variance of individual components. Once we have the covariance matrix, it is possible to compute eigenvalues and eigenvectors (principal components). A transformation matrix (TM) is formed with eigenvectors as its row. For data reduction, eigenvectors with low eigenvalue are removed from the transformation matrix. Thus, PCA projects data into eigenvector space with fewer dimensions. The transformation matrix is defined as (2). The eigenvectors are arranged in descending order within the transformation matrix. In the current work, a total of 1, 96,608 features exists and the number of principal components selected manually is 50. These principal components are capable of collecting crucial information necessary for the classification purpose. Reduced number of principal components affects the classification accuracy.

$$TM = \begin{matrix} e1 \\ e2 \\ \vdots \end{matrix} \qquad (2)$$

C. *Classifiers*

In a machine learning framework, features are given as input to the ML models which in turn extract the hidden patterns. These patterns are used to identify ‘labels’ that successively classify information. The classifiers used in the present work are Support Vector Machine (SVM) and Random Forest classifier. SVM is a supervised learning algorithm whose objective is to find a hyper plane that classifies data in an N-Dimensional space. Random Forest is based on ensemble learning where certain number of decision trees make predictions and the final supposition is based on majority votes of predictions. In the proposed method, the number of decision trees (n_estimators) in the forest is 50. Scikit-Learn library of Python is used for executing different classifiers.

IV. DATASET

Magnetic Resonance Imaging is a structured technique to display brain tissues. The controlled magnets of MRI scanner aligns the protons present in every cell to the direction of magnetic field. When the radio frequency pulse is removed, the protons returns back to equilibrium state and the response action is captured as MRI.

The dataset used in the proposed method is CE-MRI. This dataset contains 3064 images in .mat file format. In Python, these files can be read using h5py package. The dataset contains three classes of tumors i.e. Meningioma, Pituitary and Glioma from 233 patients. There are 708 images with meningioma tumor, 1426 images with Glioma tumors and 930 images with pituitary tumors. All images are T1 weighted and includes axial, sagittal and coronal views. The size of each image is 512x512. To speed up computation time, these images were resized to 256x256.

Each .mat file consists of a) image with tumor b) tumor mask for segmentation purpose c) patient Id d) labels that denote type of tumor e) coordinates of tumor in the image. The labels for Meningioma, Glioma and Pituitary are 1, 2 and 3 respectively. Fig. (4) shows an MRI with meningioma tumor. Fig. (5) and Fig. (6) depicts the corresponding tumor mask and image details.

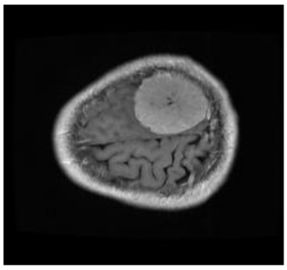


Fig.4 MRI with Meningioma tumor

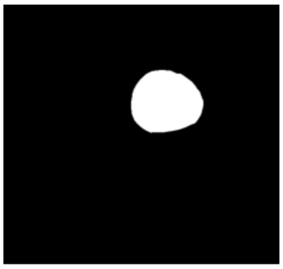
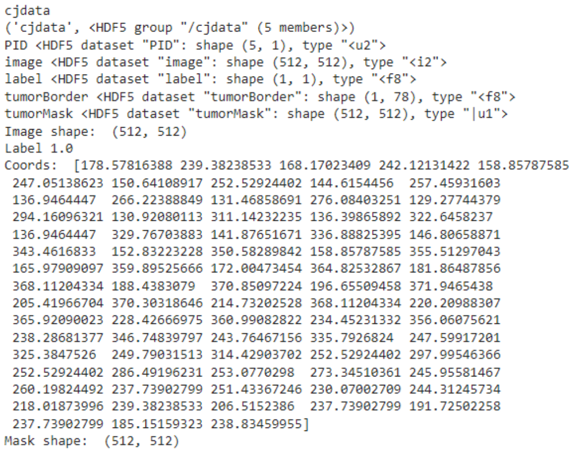


Fig.5 Tumor Mask



```
cjdata
('cjdata', <HDF5 group "/cjdata" (5 members)>)
PID <HDF5 dataset "PID": shape (5, 1), type "<u2">
image <HDF5 dataset "image": shape (512, 512), type "<i2">
label <HDF5 dataset "label": shape (1, 1), type "<f8">
tumorBorder <HDF5 dataset "tumorBorder": shape (1, 78), type "<f8">
tumorMask <HDF5 dataset "tumorMask": shape (512, 512), type "|u1">
Image shape:  (512, 512)
Label 1.0
Coords:  [178.57816388 239.38238533 168.17023409 242.12131422 158.85787585
 247.05138623 150.64108917 252.52924402 144.6154456  257.45931603
 136.9464447  266.22388849 131.46858691 276.08403251 129.27744379
 294.16096321 130.92080113 311.14232235 136.39865892 322.6458237
 136.9464447  329.76703883 141.87651671 336.88825395 146.80658871
 343.4616833  152.83223228 350.58289842 158.85787585 355.51297043
 165.97909097 359.89525666 172.00473454 364.82532867 181.86487856
 368.11204334 188.4383079  370.85097224 196.65509458 371.9465438
 205.41966704 370.30318646 214.73202528 368.11204334 220.20988307
 365.92090023 228.42666975 360.99082822 234.45231332 356.06075621
 238.28681377 346.74839797 243.76467156 335.7926824  247.59917201
 325.3847526  249.79031513 314.42903702 252.52924402 297.99546366
 252.52924402 286.49196231 253.0770298  273.34510361 245.95581467
 260.19824492 237.73902799 251.43367246 230.07002709 244.31245734
 218.01873996 239.38238533 206.5152386  237.73902799 191.72502258
 237.73902799 185.15159323 238.83459955]
Mask shape:  (512, 512)
```

Fig.6 Image details in .mat file format

V. Evaluation Metrics

The performance of the model is measured in terms of Accuracy, Precision, Recall and F1 score. The terms are defined as (2), (3), (4), (5) respectively. TP (True Positive) represents the cases in which actual output and predicted output are positive, TN (True Negative) denotes the cases when both actual output and predicted output are negative, FN (False Negative) shows the cases in which the actual output is true and predicted output is incorrect, FP (False Positive) represents the cases where the prediction is true and actual output is incorrect.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$F1\ score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{5}$$

VI. EXPERIMENTAL RESULTS

In this section, the results obtained using the proposed method is presented. The experiments in this work use Google Colab environment. The whole data is divided into training and testing data in the ratio 80:20. There are 2451 train data samples and 613 test data samples. Selecting the number of principal components is tricky and is set manually only. When n_components=2, the classification accuracy for SVM is only 70% and for Random Forest it is 68%. As we increase the component number to 50, the accuracy got improved.

Table I and Table II shows the classification report of SVM and Random Forest respectively. Support Vector Machine achieves a classification accuracy of 90% which is 2% more than Random Forest. However, the precision value of meningioma MR slices is low in both classifiers. This is due to lowered number of samples present in the dataset. The confusion matrix for SVM and Random Forest are given as Fig. (7) and Fig. (8) respectively.

TABLE I. Classification Report Of SVM

Tumor type	Precision	Recall	F1-score	Support
Meningioma	0.69	0.86	0.77	116
Glioma	0.95	0.88	0.92	313
Pituitary	0.97	0.95	0.96	184
Accuracy	**90.0%**			
Execution time	**0.27 sec**			

TABLE II. Classification Report Of Random Forest

Tumor type	Precision	Recall	F1-score	Support
Meningioma	0.63	0.85	0.73	107
Glioma	0.93	0.85	0.89	317
Pituitary	0.98	0.94	0.96	189
Accuracy	**87.6%**			
Execution time	**0.73 sec**			

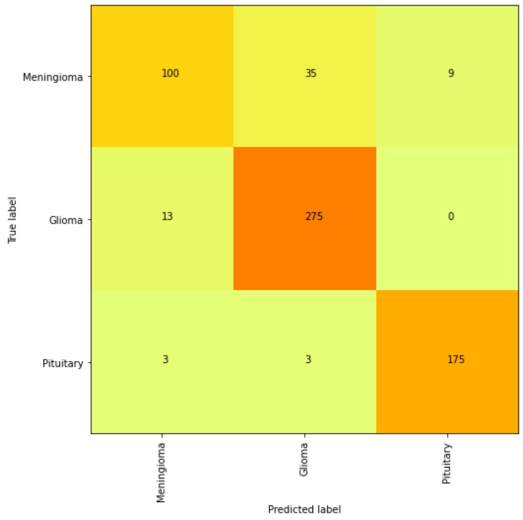


Fig.7 Confusion matrix of SVM

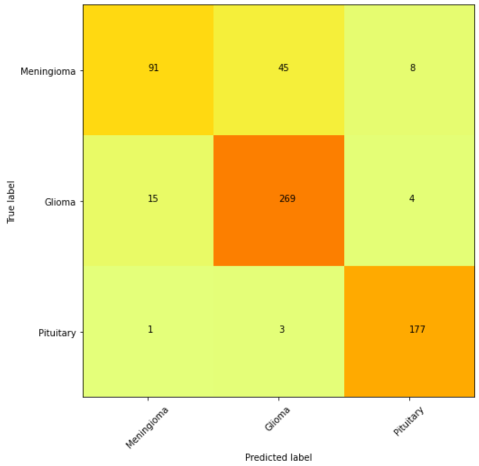


Fig.8 Confusion matrix of Random Forest

VII. CONCLUSION

In this paper, we proposed a framework for brain tumor classification based on dimension reduction. The feature size is huge and hence we use PCA for improving computation time. The SVM classifier achieves 90% accuracy in the classification task. In future, we will add normal brain images to the dataset which may differentiate classification to next stage.

ACKNOWLEDGMENT

I would like to thank my guide Prof. Dr. Robert A Singh and the faculties of School of Computing, Kalasalingam Academy of Research and Education for their advice and continued support.

REFERENCES

[1] Byale, H., Lingaraju, G.M. and Sivasubramanian, S., 2018. Automatic segmentation and classification of brain tumor using machine learning techniques. International Journal of Applied Engineering Research, 13(14), pp.11686-11692.

[2] Johnson, D.R., Guerin, J.B., Giannini, C., Morris, J.M., Eckel, L.J. and Kaufmann, T.J., 2017. 2016 updates to the WHO brain tumor classification system: what the radiologist needs to know. Radiographics, 37(7), pp.2164-2180.

[3] Vijayarajeswari, R., P. Parthasarathy, S. Vivekanandan, and A. Alavudeen Basha. "Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform." Measurement 146 (2019): 800-805.

[4] Alam, Janee, Sabrina Alam, and Alamgir Hossan. "Multi-stage lung cancer detection and prediction using multi-class svm classifie." In 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018, pp. 1-4.

[5] Dai, Bin, Rung-Ching Chen, Shun-Zhi Zhu, and Wei-Wei Zhang. "Using random forest algorithm for breast cancer diagnosis." In 2018 International Symposium on Computer, Consumer and Control (IS3C), pp. 449-452. IEEE, 2018.

[6] Swati, Z.N.K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S. and Lu, J., 2019. Brain tumor classification for MR images using transfer learning and fine-tuning. Computerized Medical Imaging and Graphics, 75, pp.34-46.

[7] Shree, N. Varuna, and T. N. R. Kumar. "Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network." Brain informatics 5, no. 1 (2018): 23-30.

[8] Hussain, Ashfaq, and Ajay Khunteta. "Semantic Segmentation of Brain Tumor from MRI Images and SVM Classification using GLCM Features." In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 38-43. IEEE, 2020.

[9] Arora, Ayush, Priyanshu Roy, S. Venktesan, and Ramesh Babu. "K-NN based classification of brain MRI images using DWT and PCA to detect different types of brain tumour." International Journal of Medical Research & Health Sciences 6, no. 9 (2017): 15-20.

[10] https://figshare.com/articles/dataset/brain_tumor_dataset/1512427

[11] Hussain S, Anwar SM, Majid M. "Brain tumor segmentation using cascaded deep convolutional neural network". Annu Int Conf IEEE Eng Med Biol Soc. 2017 Jul;2017:1998-2001. doi: 10.1109/EMBC.2017.8037243. PMID: 29060287.

[12] Mirzaei, Fazel, Mohammad Reza Parishan, Mohammadjavad Faridafshin, Reza Faghihi, and Sedigheh Sina. "Automated brain tumor segmentation in MR images using a hidden Markov classifier framework trained by SVD-derived features." Image Video Process 9 (2018).

[13] Ghassemi, Navid, Afshin Shoeibi, and Modjtaba Rouhani. "Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images." Biomedical Signal Processing and Control 57 (2020): 101678

[14] Hebli, Amruta, and Sudha Gupta. "Brain tumor prediction and classification using support vector machine." In 2017 international conference on advances in computing, communication and control (ICAC3), pp. 1-6. IEEE, 2017.

[15] Krishnakumar S, Manivannan K et al "Effective segmentation and classification of brain tumor using rough K means algorithm and multi kernel SVM in MR images" J Ambient Intell Human Comput 12, 2021, pp 6751–6760

[16] Mudgal, Tushar Kant, Aditya Gupta, Siddhant Jain, and Kunal Gusain. "Automated system for Brain tumour detection and classification using eXtreme Gradient Boosted decision trees." In 2017 International Conference on Soft Computing and its Engineering Applications (icSoftComp), pp. 1-6. IEEE, 2017.

[17] Anitha, R., and D. Siva Sundhara Raja. "Development of computer-aided approach for brain tumor detection using random forest classifier." International Journal of Imaging Systems and Technology 28, no. 1 (2018): 48-53.