Journal of Pharmaceutical Research International

33(59A): 790-797, 2021; Article no.JPRI.79426
ISSN: 2456-9119
(Past name: British Journal of Pharmaceutical Research, Past ISSN: 2231-2919, NLM ID: 101631759)

Brain Tumor Classification using Machine Learning

N. Pavitha [a], Atharva Bakde [b], Shantanu Avhad [b], Isha Korate [b], Shaunak Mahajan [b] and Rudraksha Padole [b*]

[a] Department of Artificial Intelligence and Data Science, Vishwakarma Institute of Technology, Pune, Maharashtra, India.
[b] Department of Information Technology, Vishwakarma Institute of Technology, Pune, Maharashtra, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JPRI/2021/v33i59A34330

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/79426

Original Research Article

Received 10 October 2021
Accepted 14 December 2021
Published 16 December 2021

ABSTRACT

This paper presents a technical analysis of tumor data with Machine Learning and Classification Approach. Feature parameters which are dependent for classification of tumor are used for analyzing and classifying the class of tumor. In the classification of tumor, KNN-Classifier is implemented with cross validating accuracy score and tuning hyper parameters. Experimental simulation for best average score for K makes it to the cross validation. Approaching the prediction with the best accuracy score, hyper parameters of KNN Classifier states the best score. Using Principal Component Analysis on the data, miss-classification of tumor class in data is visualized.
Aims: To declare and analyse tumor data from the source of MRI, CT scan, etc. for medication of tumor. To utilize smart predictions for the upcoming tumor patients using Machine Learning.
Study Design: Tumor classification using K Nearest Neighbor algorithm and analysis of the miss-classification.
Methodology: We included 11 different studies and research papers which were relevant with tumor classification. Research papers include classification of tumors with different supervised learning approaches. Our proposed analysis and classification give visualization of two classes of tumor.
Results: The Project results in classification of tumor data using Machine Learning and analyzing the miss-classification of tumor. In implementation of KNN Algorithm, the accuracy score after

**Corresponding author: E-mail: rudraksha.padole19@vit.edu;*

cross validation and tuning K values is 0.97. The confusion matrix shows 4 false positives and 1 false negative value in testing.
Conclusion: Less miss-classification of tumor results best accuracy score and more efficient working on testing data. Visualizing the classification with 3-dimensional scatter plots made the analysis accurate.

Keywords: Data analytics; Machine Learning (ML); classification; tumor classification; K-nearest neighbor; principal component analysis.

1. INTRODUCTION

Cancer is the one of the prime reasons behind loss of life all-over the world. In step with the World Health Organization (WHO) early detection of cancer can forestall death, however this cannot be invariably potential. In contrast to cancer, a growth may be benign, pre-carcinoma or malign. Benign Tumors are different from malignant tumors in a way that they don't spread to alternative organs and tissues. Also, doctors are often able to remove them surgically.

The most common methodology for differential medicinal growth of this kind is MRI short for Magnetic Resonance Imaging. However, early detection of tumor exceptionally relies on the expertise of the radiologists. Before it has been established that whether the tumor tissue is malignant or benign the diagnostics won't be completed. In today's world where there are new technologies and new innovations happening each and every day, it is really vital to develop a reliable and efficient medicine tool/ technology for tumor growth subdivision and classification of MRI images [1].

In order to ensure early Categorization and Detection of Brain Tumor developing a system is substantially important. Since wrong identification gives rise to malignant results, the classification techniques ought to give high level accuracy while carrying out Tumor classification. High accuracy with ML algorithms and techniques used for detection and classification is often obtained with the right approach of algorithm on data. Implementation of the correct classification algorithm on data leads to the best accuracy of algorithms. Different parameters are needed to classify the class of tumor. These parameters are then combined as they are dependent for classification. Daily variety of tumor patients are in labs using MRI/CT scans and also the features of tumors are stored as the data of patients. This data can then be used for analysis and prediction of other patients. In technical terms Machine Learning can be the solution to analyze and predict the class of tumor. This can help the Radiologists, Neurologist and Neurosurgeon as they ease work through the support of computer-based systems in developing New Techniques for Health care related issues [2].

This research paper is divided into six sections. The first section is Introduction, which gives a gist about why there's a need for a classification and detection system. Also, it states the difference between Benign tumor and Malign tumor. The second section includes the literature survey of different papers in this domain. The next in enumeration comes the Implementations Details section. In this section, the tools used for the proposed methodology for classification of tumor are mentioned. The workflow of the whole study is demonstrated using a flow diagram. In addition to this, the details of how the model is implemented is mentioned. Results and Discussion is the fourth section. In this section the result i.e., the accuracy score is mentioned, heat map and scatter plots are included to exhibit the outcome in a more descriptive aspect. Fifth section is Conclusion and Future scope. This section has concluded the research and mentioned some points which can be researched about. The last sections comprise of all the references studied and the link for the dataset used in this study.

2. LITERATURE REVIEW

Brain Tumor comes under the rapid dreadful diseases' difficult ample folks. Tumor Detection and organization is important to confirm early Detection and Categorization of the tumor. Since Wrong identification ends up in Dreadful and Fatal results, the classification and segmentation techniques ought to offer High-Level Accuracy whereas activity tumor classification. These Techniques ought to be effective for period applications. The Growing urge for saving Time is equally vital because of the Desired Accuracy Level of the Result.

The want for a Full-fledged analysis on Comparative base marks its Significance this will profit Radiologists as they ease to go through the

support of computer systems in developing new techniques for peaked Health care issues.

S.Nagini, B.V.Kiranmayee, Dr. TV. Rajinikanth in 2016 used an associate approach which had a coaching and testing stage to find out the tumor. The practicality of the planned algorithmic program has been valid by constructing a paradigm Application [2].

Viktor Losing, Barbara Hammer, and HeikoWersing in 2016 counsel SAM short for SELF ADJUSTING MEMORY show for the KNN algorithm program. Since KNN builds up a confirmed classifier [3].

Dr. M.A. Ansari, Garima Singh [4] planned a completely distinctive methodology that has normalization of K-means Segmentation and bar graph.

Dr. K. Satay Ras K. Sudha ani, Dr. T. C. Samra [5] designed a methodology that embodies ways like a bar graph, distance matrix, re-sampling, K-NN algorithmic program. Identification and classification of tumor by k-NN that relies on training of k. Manhattan metric did the calculation and application of the classifier during this time period.

Single Image Super-Resolution is applied on tomography to extend the resolution. options are extracted from the pre-trained ResNet design. SVM short for Support Vector Machine, Binary classification provides ninety-fifth accuracy Mean shift bunch technique is employed for segmentation in tumor classification exploitation Edge adjustive Total Variation [6].

Feedforward neural network is tested on AANLIB dataset, a dataset published by the renowned Harvard Medical School containing 239 pictures with ninety-nine accuracy exploitation Haralick texture options. Multigrade brain organization Exploitation CNN short for Convolutional Neural Network is planned in [7].

This paper presents the classification of tumors exploitation Convolution Neural network, Al and deep learning's vital advancement created medical image process technique a lot economical, and that helps the doctor to diagnose sickness simply. Here with the assistance of tomography pictures the tumor is assessed into cancerous and non-cancerous. Here the methodology is initial: the info set is delineated (MRI pictures tumor knowledge set), then the info is preprocessed with facilitation of ecstasy pooling flattening full affiliation various techniques (Convolution operation, liquid ecstasy pooling, flattening, full affiliation). Further machine learning is employed to classify the dataset here with facilitate of some algorithms Here the target is to classify the tomography pictures of brain tumors, as we all know convolution neural network classification is employed, it's wont to sight pictures exploitation Keras by developing a convolution neural network, here initial exploitation image detection the world of the region of interest of facilitating of a strategy for tumor classification by tomography pictures is employed, then with proposing a straightforward CNN network And to induce correct results neural network wants massive knowledge to coach on, and more the information is trained and also classified properly.

Deep learning and Machine learning [8]. Ali Ari et al. have reported a paper within which they need to use Deep Learning for tumor classification detection. It depends on the brain on cancer, the treatment physician's experience and data, for this reason, victimization associate degree automatic growth finding technology is of paramount importance to sight brain tumors. The projected methodology is divided into three phases, the area unit pre-processing, the acute learning machine native receptive fields (ELM-LRF), and image method. At first, the research performs native smoothing methods that were accustomed to exclude the possible noises. Among the second stage, by using victimization ELM-LRF, the adult male footage was classified as benign or malignant. In the third stage, the tumors were further subdivided, the objective of the study was victimization alone adult male footage, that has a mass, therefore avoiding wasting the physician's time. Among the preliminary studies, the classification accuracy of adult male footage comes out to be ninety-seven. Graded results revealed that the presented methodology's performance was surpassing the alternative recent studies among the literature survey. Experimental outcomes jointly verified that the presented methodology is constructive and could be applied in portable computer power-assisted tumor detection [9].

To come up with a brand-new CNN architecture for classifying brain tumors while working on brain MRI images was the main aim of this proposed model. First of all, the dataset that is described here is a contrast Enhanced brain MRI's images and T1 weighted, constituting 3

different tumors. So further in the methodology, Convolution neural network is used here. CNN consists of many layers that can treat many kinds of data inputs. It uses two processes for training and testing. The database is here split into training which is 70%, testing part 30%.

Furthermore, this research explains the proposed CNN architecture , here this architecture consists of 18 layers to classify the brain tumors more efficiently. This technique is applied on 3 different datasets namely cropped, uncropped and segmented. CNN architecture grades the brain tumor in the above cases and makes a confusion matrix for all cases. 4 statistical indices are calculated with the help of the created confusion matrix which are accuracy, sensitivity, specificity, and precision.

Further the dataset was divided into the training part which constituted 70% of the overall data, validation part included 15% and testing also had 15%.

Further the accuracy for(64x 64) uncropped image was (99.0%), cropped image (98.4%), and segmented image is (97.6%) [10].

This proposed paper classifies intracranial tumor from MRI using wavelets and ML. Here the data from multi-modal intracranial tumor subdivision challenge is utilized.

First of all, in this study they pre-process the Brain Tumor dataset which has four modalities of MRI. All the images are visualized and to enhance the image contrast they have performed histogram matching. The next step is to determine the bounding box around the tumor area.

Further, feature extraction is performed, it has 4 features namely, neighborhood information, intensity, intensity difference, and wavelet-based textures. From the images these features are extracted, then these images are used to perform supervised classification.

Further, the supervised algorithms like KNN, Random Forest, AdaBoostM2 and Random Under Sampling (Rus Boost) are compared here. As a result of which random forest algorithm is chosen, since it has become a significant data analysis tool, also it can be applied to nonlinear and high order datasets.

In the final results section, the result of the proposed model is compared with prior works on the brain tumor dataset of the patients [11].

3. METHODOLOGY

3.1 Implementation Details

Implementation of this research is of Tumor data and processing it with Machine Learning Approach. The Tumor dataset has 699 rows and 11 columns comprising the data of patients [12]. The structure of the data frame is for classification of tumor into two classes. The tumor dataset has the predictor categorical column class having two unique tumor class values which classify the data of tumor patients.

Fig. 1. Workflow of the project

Brain tumor is graded into four distinct grades. Grade I being the one with most survival rates and IV being the least. In the dataset considered for this study there are grade II i.e., Class 2 and grade IV i.e., Class 4 unique values for classification of Tumor. Depending on 9 parameters for classification of tumor are numeric and continuous values. Analysis started with cleaning the data (dropping rows having '?' values). Count of class '2' is 444 and class '4' is 241 shown by the count plot.

For classification, Machine Learning K-Nearest Neighbor Algorithm (classifier) is implemented. Separating the features and target (class) columns and splitting data into training and testing with test size = 0.3. . Value of K is often determined by taking square root of the size of data i.e., of the number of rows which for this dataset is 699 but as some rows had '?' values, they had to dropped for further implementation, square root of the remaining number of rows comes out be 21. Taking K=21, the model fits the data giving the model accuracy score and miss classification values of the testing data using a confusion matrix. Applying statistics on miss classification values, mean error value is determined. Checking the accuracy score and error value for K ranging 1 -50 through. This accuracy scores and error rates differ as per the testing data. Cross validation, an important technique to cross verify the accuracy scores taking average of accuracy scores of 10 testing data samples [13].

For building the model with the best accuracy score, KNN, short for K nearest-neighbour classifier, was trained with hyper-parameter i.e., n_neighbours. For the value of n 46 the model gives the best accuracy. With this value the model fits again giving the final accuracy score and confusion matrix for KNN Classifier. Predictions of the Tumor class were compared with the actual class which gave the final report. PCA which is short for Principal component analysis is used to transform the features into explained variance and components. This is used for 2D and 3D plotly visualisation of the classification of Tumour class. Tumour data is transformed in 2 components i.e., PC1, PC2 for 2D visualisation seaborn. For 3D visualisation data is transformed in 3 components i.e., PC1, PC2, PC3 having the explained variance. After visualisation in Plotly Dash, the classification of tumour data can be analysed.

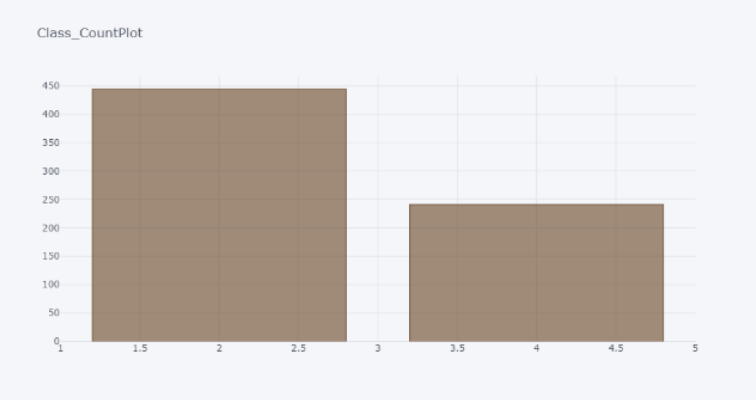

Fig. 2. Count Plot of Tumor Class

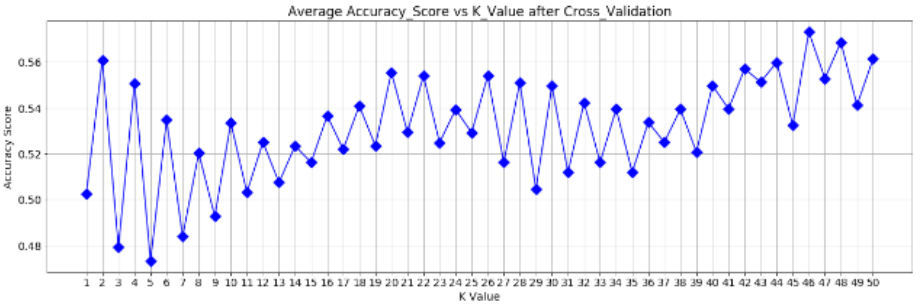

Fig. 3. Line graph of average accuracy score with K- values after cross validation

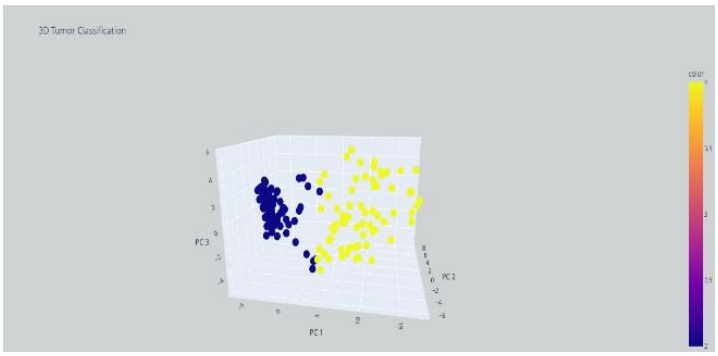

Fig. 4. Plotly Dash visualizing 3D scatter plot

4. RESULTS AND DISCUSSION

The research results in classification of tumor data using Machine Learning and analyzing the miss-classification of tumor. In implementation of KNN Algorithm, the accuracy score after cross validation and tuning K values is 0.97.

Confusion matrix is usually used when there is binary classification involved. Confusion matrix compares the predicted labels to the true labels. The true negative values came out to be 134, on the other hand false negatives were 4. True positives were 66 and false positives was 1. The heat-map of the confusion matrix with very less miss classified values is shown below.

Visualization of actual and predicted class is done with Principal component analysis (PCA).

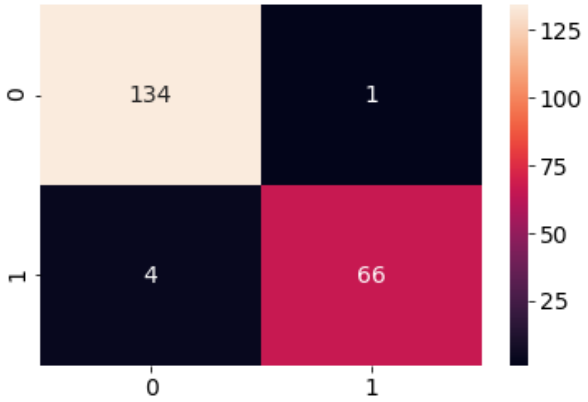

Fig. 5 Heat map of confusion matrix showing classification

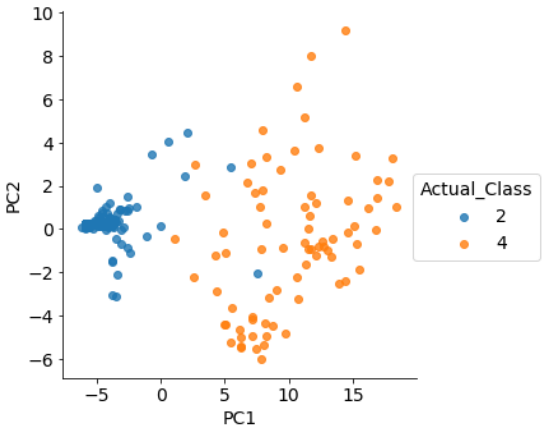

Fig. 6. Scatter plot of actual values (class) using PCA

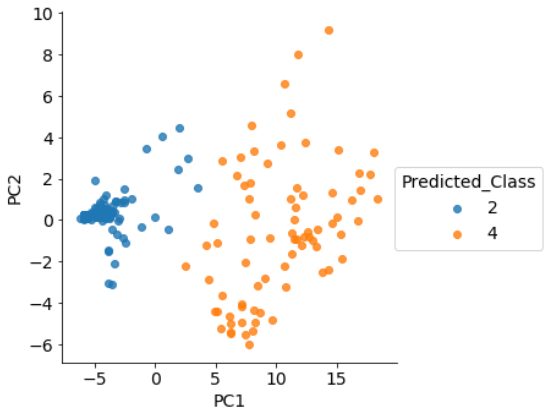

Fig. 7. Scatter plot of predicted values (class) using PCA'

For predictive modeling and exploratory data analysis Principal Component Analysis was used in this research. The scatter plot in Fig. 6 and Fig. 7, the actual values, and the predicted values respectively came out to be almost similar, which shows that the trained model performed well on the test data.

5. CONCLUSION

Evaluating the data with Machine Learning algorithm, classification of tumor is implemented and the accuracy score in testing is 97% of KNN Classifier. As per the topic, the classification according to the features must be visualised to analyse the miss classification. With this system and workflow for data, the objectives like accurate predictions and visualisation of data are completed. Future scope would be, implementation of ANN, CNN, RNN, etc. DL algorithms after DL approach with TensorFlow. Implementation of MRI images and classification of tumor with this approach can be done. An interface for more specific classification can be made with real time parameters affecting tumor can be implemented. Future scope would be implementation of Deep Learning algorithms after estimating the data using Tensor Flow. Implementation of MRI images and classification of tumor with this approach can be done. An interface for more specific classification can be made with real time parameters affecting tumor can be implemented.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Milica M, Badža , Marko C, Barjaktarovi´C. Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network. MDPI, MultiDisciplinary Digital Publishing institute; 2020.
2. Kiranmayee BV, Rajinikanth TV, Nagini S. A Novel Data Mining Approach for Brain Tumor Detection. 2016;978-1-5090-5256-1/16/$31.00_c IEEE
3. Viktor Losing, Barbara Hammerand HeikoWersing. KNN Classifier with Self Adjusting Memory for Heterogeneous Concept Drift", 2016 IEEE 16th International Conference on Data Mining; 2016.
4. Garima Singh DR, Ansari MA. Efficient Detection of Brain Tumor from MRIs Using K-Means Segmentation and Normalized Histogram. In IEEE Issue; 2016.
5. Sudharani K, Sarma TC, Satya Rasad K. Intelligent Brain Tumor Lesion Classification and Identification from MRI Images Using k-NN Technique. In International Conference on Control, Instrumentation, Communication.
6. Vallabhaneni RB, Rajesh V. Brain tumour detection using mean shift clustering and GLCM features with edge adaptive total variation denoising technique. Alexandria Engineering Journal. 2018;2387-2392.
7. Sajjad M, Khan S, Muhammad K, Wu W, Ullah A, Baik SW. Multi-grade brain tumor classification using deep CNN with

extensive data augmentation. Journal of Computational Science. 2018;174-182.
8. Saranya N, Karthika Renuka D, Nikkesh Kanthan J. Brain Tumor Classification using Convolution Neural Nework. Journal of Physics. 2021;1742-6596.
9. Ali Ari, Davut Hanbay. Deep learning based brain tumor classification and detection system", Turkish journal of Electrical Engineering and Computer Sciences, 2018.a
10. Ali Mohammad Alqudah, Hiam Alquraan , Isam Abu Qasmieh , Amin Alqudah , Wafaa Al-Sharu, "Brain Tumor Classification Using Deep Learning Technique - A Comparison between Cropped, Uncropped, and Segmented Lesion Images with Different Sizes" in International Journal of Advanced Trends in Computer Science and Engineering. 2019;8(6).
11. Usman Khalid, Rajpoot Kashif. Brain tumor classification from multi-modality MRI using wavelets and machine learning, Pattern Analysis and Applications. 2017;20:871-881.
12. Verma G, Verma H. Hybrid Deep Learning Model for Emotion Recognition using Facial Expressions", The review of socionetwork strategies. Springer (ESCI INDEX). 2020;14:171-180.
ISSN – 1867-3236)
DOI: 10.1007/s12626-020-00061-6
13. Verma G, Verma H. A Multilayer Perceptron Neural Network Model for Predicting Diabetes", International Journal of Grid and Distributed Computing. (ESCI, Scopus Index, H-Index-6). 2020;13(1): 1018-1025.
ISSN-2005-4262
DOI: http://sersc.org/journals/index.php/IJGDC/article/view/17405.

Reference to Web-resource or Electronic Articles

[1] Dataset:
https://archive.ics.uci.edu/ml/index.php

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/79426