



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 2, February 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

A Comparative Analysis of Convolution Neural Network and Machine Learning Algorithms in Brain Tumor Prediction

Chethan C¹, Pankaja R², Raghavendra S N³

Assistant Professor, Department of Information Science & Engineering, Sri Venkateshwara College of Engineering,
Bangalore, Karnataka, India ¹⁻³

ABSTRACT:In the recent times, detection of brain tumor has turned up as a general causality in the realm of health care. Brain tumor in very simple words can be described as a malformed mass of tissue wherein the cells multiply abruptly and ceaselessly and there is no control over the growth of the cells. Deep Learning (DL) can be used to repeatedly to categorize images and objects with almost human-level precision with its ability to perform unsupervised feature extraction over massive data sets making big data parts of the solution not the problem. This paper proposes an automatic segmentation method that relies upon CNN (Convolution Neural Networks), determining small 3×3 kernels. CNN (a DL technique) from NN (Neural Networks) wherein it has layer based for results classification. The results obtained by CNN are then compared with the other classical machine learning classifiers like Naïve Bayes, Random Forest and Support Vector Machine. The techniques of CNN and Machine Learning are effectually implemented in brain tumor detection and prevention.

KEYWORDS:Brain Tumor, Convolutional Neural Networks, Naïve Bayes, Random Forest, Support Vector Machine

I. INTRODUCTION

Tumor basically symbolizes abnormal and uncontrollable growth of cells within the body. Brain tumor signifies a malformed mass of tissue wherein the cells multiply abruptly and ceaselessly within the brain tissues [1]. Machine learning is a part of Artificial Intelligence (AI) that delivers systems the capability to automatically learn and improve from the experience without the need to be explicitly programmed. [2] Deep learning on the other hand is a subdivision of Machine Learning in Artificial Intelligence (AI) that has the networks proficient in learning in an unsupervised form from data that is unstructured or unlabeled[3]. It is an artificial intelligence technique that emulates the workings of the human brain in processing of data and creating patterns to be used in decision making process [4]. A Neural Network is a sequence of algorithms that provides in the recognition of the underlying relationships from a set of data through a process that imitates the way the human brain will operate [5]. It is one of the classification technique based on Bayes' Theorem. Naive Bayes classifier comes with an assumption that the presence of a feature in a class is unrelated to the presence of any other feature.

Support Vector Machine (SVM) is used for both classification and regression challenges. It is a supervised Machine Learning algorithm where each data item is plotted as a point in n-dimensional space (n: number of features) with the value of each feature being the value of a particular coordinate. A Convolutional Neural Network (ConvNet/CNN) is Deep Learning algorithm which takes an input image, assigns importance (learnable weights and biases) to different aspects/objects in the image to be able to differentiate one from the other [6] [7].

The paper focuses on to how build a detection system related to brain tumor by incorporating machine learning algorithms such as Support Vector Machine(SVM), Naïve Bayes, Random Forest and Convolutional Neural Networks which provides a Deep Learning Approach and provides higher accuracy compared to the Machine Learning Algorithms.

II. EXISTING SYSTEM

Different magnetic resonance imaging (MRI) sequence images are employed for diagnosis, including T1-weighted MRI, T2-weighted MRI, fluid-attenuated inversion recovery MRI, and proton density-weighted MRI [8]. The recognition of a brain tumor at an initial stage is a key thing in providing efficient treatment. Once a tumor is clinically alleged, radiological assessment is needed to examine its size, location and impact on the surrounding tissues. On the

basis of this information the best therapy, surgery, radiation or chemotherapy is decided. Convolutional Neural Networks involve image feature analyzing to perform tasks such as segmenting tumors. This mainly includes training the network with a manually segmented dataset which is later poised to segment patient images.

In the paper [9] the accuracy of brain tumor detection model by using ANN and CNN is compared. The accuracy we got by using basic ANN is 89.6%, precision is 90.4% , sensitivity is 88.5%, specificity is 90.8% as in ANN we only use one layer, we can't apply ANN for more than one layers which results in less accuracy ,precision, sensitivity , specificity values. CNN is used for 3 layers and CNN with un-augmented dataset got the accuracy as 92.3%, precision is 95.2%, sensitivity is 93.4%, specificity is 91.8% and CNN with augmented dataset the accuracy is 94.1%, precision is 96%, sensitivity is 95.5% and specificity is 93.2%

The paper[10] implements six classifiers, namely Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Multilayer Perceptron (MLP), Logistic Regression, Naïve Bayes and Random Forest and then moves onto Convolutional Neural Network (CNN) which is implemented using Keras and Tensorflow and the results show that it yields a comparatively better performance than the six classifiers with an accuracy of 97.87%.

III. PROPOSED SYSTEM

The objective of this paper is to predict the occurrence of Brain Tumor. In the process of predicting the success we have implemented three different machine learning algorithms and one Deep learning algorithm. We have considered some of the parameters to check the efficiency of a machine learning algorithm.

1. Gathering data

The Brain Tumor is a database which is open source includes MRI Scans in the brain. The number of cases included is around 250 including positive and negative cases of Brain Tumor. The data source is Kaggle. It has been considered the most comprehensive dataset on Brain Tumors in the world.

2. Preparing data

The initial process is preparing the data by cleaning it and transforming the raw data before processing and analysis. It often includes reformatting data, making alterations to data and the merging of data sets to enhance data.

Next, we do Feature Extraction from the Images. We employ Gaussian Blurring, Thresholding and Contours to isolate the tumor based on the color intensity. Then the features like Area, Diameter, Eccentricity, Major Axis, Minor Axis, Convex Hull etc. are extracted from the contour boxes. We prepare them in a list, make a data frame and then export it as a CSV File. This file is then later loaded, and the Machine Learning Algorithms are trained on this feature extracted dataset.

3. Choosing model

We have chosen three different machine learning algorithms here - Random Forest, Support Vector Machine (SVM), Naïve Bayes.

3.1.1. Random Forest

Random forest solves both classification and regression type problems with a decent estimation level. One of the major advantages of Random Forest is the ability to handle huge data sets with higher dimensionality. It can handle thousands of input variables and identify most significant variables so it is considered as one of the dimensionality reduction method. Further, the model outputs a very prominent thing of showcasing importance of variable. It has an effectual method for approximating missing data and upholds accuracy when large quantity of the data is missing.

3.1.2. Support Vector Machine

SVM can efficiently handle non-linear data. SVM has L2 Regularization feature making it good in generalization capabilities which prevents it from over-fitting. SVM can be utilized to resolve both regression and classification problems. SVR (Support Vector Regression) is used for regression problems and SVM is used for classification problems.

3.1.3. Naïve Bayes

When we assume the value as true for assumption of independent predictors, a Naïve Bayes classifier executes better when compared to with models. Naïve Bayes requires trivial quantity of training data to assess the test data making training period is comparatively minor.

4. Training

The training of Machine Learning model encompasses having an ML algorithm to learn from the training data. The training data must comprise of target or target attributes which represents the correct answer. The learning algorithm will catch patterns in the training data that will map the input data attributes to the target and outputs an ML model that captures these patterns.

5. Prediction

Predictive modeling helps in a building a model that can make predictions. This comprises of a machine learning algorithm that makes predictions from a training dataset. Predictive modeling can be either Regression or pattern classification.

6. Evaluation of the model

The objective of model evaluation is to approximate the generalization accuracy of a model on future (unseen/out-of-sample) data. The choice of model evaluation metrics required to quantify model performance depends on the particular machine learning task like regression classification, clustering , ranking, topic modeling.

IV. IMPLEMENTATION

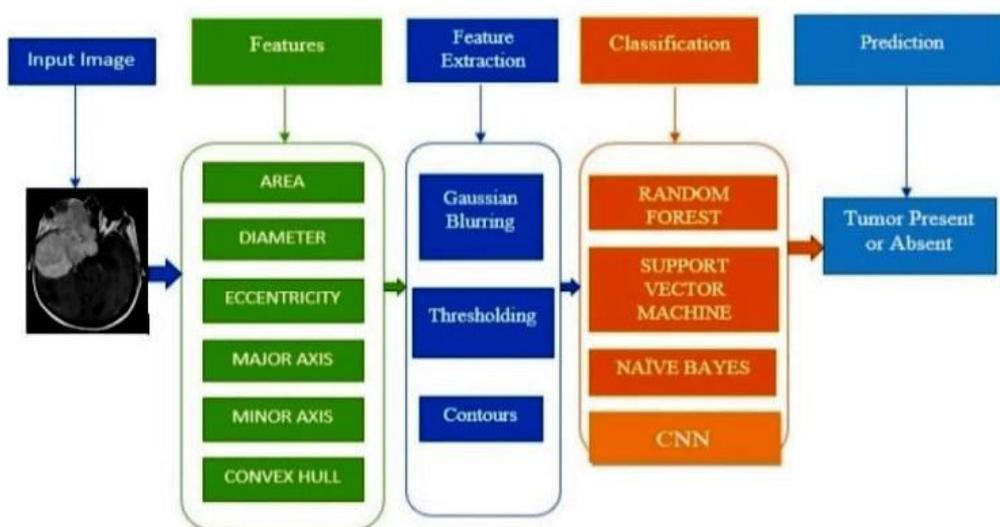


Figure 1:System Architecture

The system architecture describes the entire process as shown in the fig 1. Firstly, the brain tumor MRI (magnetic resonance imaging) is converted to grayscale image, because contours image work only on grayscale image. We apply gaussian blurring to get the object of the brain tumor part. features like area, convex area, perimeter, convex area, major axis, minor axis, eccentricity, solidity is extracted and stored in the data Frames.

The data Frames are exported to .CSV file for Classification. classification models like support vector machine (SVM), Naïve Bayes (NB), Random forest Classifier (RFC). Classifier Models extract the data and process the data and then the data is split into training and testing data. Finally, the classifier generates the confusion matrix with the accuracy and the real time classifier predicts whether the brain tumor is present or absent.

At the first MRI Scans of Brain as image data is taken. Then data is scaled and prepared for further process, data is prepared for further preprocessing. From this data the listed features are extracted and then the data is divided into testing and training data sets. These testing and training datasets are combined for further prediction.

The result obtained after choosing the model and after testing it with training data is used to generate confusion matrix. The obtained confusion matrix is used for performance analysis and to evaluate the final model.

CNN image classifications take an input image, process it and will classify the image under certain categories. Technically, deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernels), Pooling, fully connected layers(FC) and apply SoftMax function to classify an object with probabilistic values between 0 and 1.CNN process an input image and classifies the objects based on values.



The first step in extraction of features from an input image is Convolution. This conserves the relationship between pixels by learning the features using squares of input data. Convolution of an image can perform edge detection, blur, and sharpen by applying filters with different filters.

We perform convolution on the image and then apply ReLU activation to the matrix. Thereafter add convolution layers until satisfied. Feed into a fully connected layer (FC layer) after flattening the output. Output the class using an activation function (logistic Regression with cost functions) and classifies images.

1. Data Set

The dataset used in this experiment is from Kaggle, a subsidiary of Google LLC. The dataset used is composed of MRI Scans of the Brain within the range from 250-300, 50% with Positive tumor and rest 50% with Negative tumor. Each column in the table represents the features that are to be analyzed and extracted from the Image Dataset, and each row corresponds to the recorded values. The main aim of the data is to discriminate healthy people from those with Brain tumor, according to class column which is set to 0 for negative and 1 for positive. The dataset taken for Parkinson’s disease for analysis is represented in the fig.2 and fig.3.

Area	Perimeter	Convex	Ai	Solidity	Equivalen	Major Axis	Minor Axis	Eccentrici	Class
0	6314.5	521.9726	9855.5	0.640708	89.66533	86.29754	146.9529	(8.439724)	1
1	52776.5	1189.82	59947	0.880386	259.2241	209.5902	405.8614	(1.015395)	1
2	50232.5	1415.349	69628.5	0.721436	252.8992	223.6327	373.0776	(8.176507)	1
3	4486	275.6224	4741.5	0.946114	75.57614	64.81838	93.0834	(6.311037)	1
4	145	92.08326	323.5	0.448223	13.58748	16.90097	28.91504	(8.500079)	1
5	4767.5	496.3158	6550	0.727863	77.91129	69.41057	107.6818	(7.262588)	1
6	32059	1447.568	57016	0.562281	202.0366	203.8815	340.5204	(8.191247)	1
7	54709	1369.235	63054.5	0.867646	263.9274	261.2904	279.041	(2.295058)	1
8	4123.5	279.865	4550.5	0.906164	72.45829	63.91366	88.793	(5.905213)	1
9	12913	1399.308	25064.5	0.515191	128.2238	128.9773	195.8319	(6.995957)	1
10	9538	2055.891	74793.5	0.127524	110.2005	269.5944	331.2054	(4.369825)	1
11	6028.5	542.3991	8285.5	0.727596	87.61121	74.48892	129.7051	(8.728587)	1
12	395	76.08326	408.5	0.966952	22.42609	22.05364	23.88254	(2.544914)	1
13	3008	313.0782	4086.5	0.736082	61.88622	49.95913	113.8035	(1.253241)	1
14	1160.5	203.5807	1674	0.69325	38.43949	27.03943	69.99866	(1.462118)	1
15	2111	388.6173	4407.5	0.478956	51.84408	65.21905	104.9911	(7.724817)	1
16	6403.5	1622.32	91829.5	0.069732	90.29501	337.6962	465.7452	(5.815940)	1
17	791	118.5685	842	0.93943	31.73535	32.0209	37.81359	(3.846116)	1
18	60164.5	938.3646	60763.5	0.990142	276.774	242.0734	319.2044	(5.263047)	1
19	59864.5	932.5656	60230	0.993932	276.083	243.4071	317.0455	(5.110569)	1

Figure 2:Dataset of Brain Tumor with Class=1

149	0	0	0	0	0	0	0	0	0
150	0	0	0	0	0	0	0	0	0
151	0	0	0	0	0	0	0	0	0
152	0	0	0	0	0	0	0	0	0
153	490	176.1665	1101	0.44505	24.97774	22.78448	58.03669	(1.434490)	0
154	1026	159.397	1232.5	0.832454	36.14338	28.60386	54.41192	(9.908644)	0
155	0	0	0	0	0	0	0	0	0
156	0	0	0	0	0	0	0	0	0
157	89.5	56.38478	123.5	0.724696	10.67497	8.409369	22.99472	(1.558364)	0
158	445.5	137.1543	758.5	0.587343	23.81655	19.56398	49.91642	(1.437314)	0
159	1306	205.3381	1937.5	0.674065	40.77807	41.41377	62.02429	(6.826856)	0
160	245	63.1127	252	0.972222	17.66193	13.29298	24.61654	(9.543859)	0
161	568	111.2548	710	0.8	26.89238	29.79774	30.87045	(1.657753)	0
162	70.5	34.38478	74	0.952703	9.474354	7.408895	13.3682	(9.196394)	0
163	2191	612.3574	6217	0.352421	52.81731	74.35757	81.11824	(2.669820)	0
164	0	0	0	0	0	0	0	0	0
165	0	0	0	0	0	0	0	0	0
166	170.5	185.3797	770.5	0.221285	14.73388	77.76366	117.5247	(6.938587)	0
167	0	0	0	0	0	0	0	0	0
168	65.5	44.04163	81.5	0.803681	9.132206	5.33636	22.18497	(2.470885)	0
169	264	130.5097	409.5	0.644689	18.33399	10.36228	56.33717	(3.272254)	0

Figure 3:Dataset of Brain Tumor with Class=0

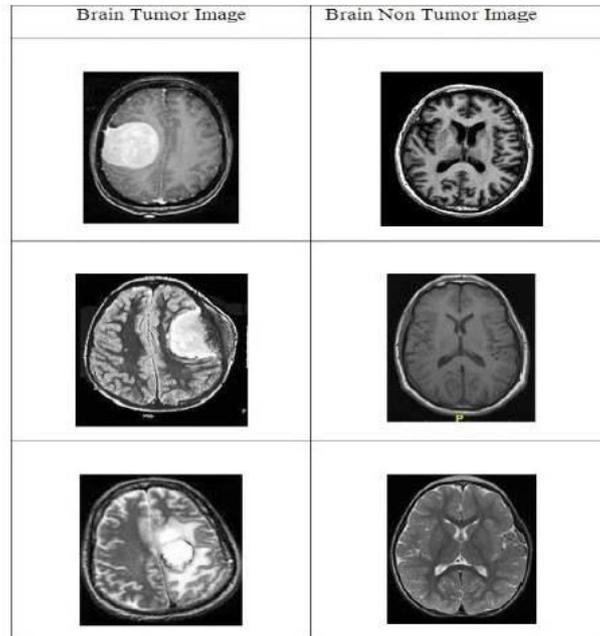


Figure 2: MRI Dataset of Brain Tumor

A total of 8 features are kept for pre-processing of the image that has been selected. Table 1 gives a brief explanation of meaning of the features.

Table 1. Brain Tumor Features

Feature	Description
Area	It denotes the number of elements in the region of interest (ROI) represented.
Perimeter	It represents the distance around the boundary of the segmented region of interest (ROI).
Convex Area	It represents the total number of elements in a convex image within the region of interest (ROI).
Solidity	It represents the proportion of the elements in the convex hull, which are also in the region of interest (ROI).
Equivalent Diameter	It measures the diameter of the circle with the same area as the region of interest (ROI)
Major Axis and Minor Axis	Major axis and minor axis length represent the length of the major and minor axis of an ellipse respectively that have same normalized second central moments as the region of interest (ROI).
Eccentricity	It specifies the eccentricity of the ellipse that has same second-order moments as the region of interest (ROI). It denotes the ratio of the distance between the foci of the ellipse and its major axis length.

V. RESULTS

By definition a confusion matrix C is such that $C_{i,j}$ is equal to the number of observations known to be in group i and predicted to be in group j. Thus in binary classification, the count of true negatives is $C_{0,0}$, false negatives is $C_{1,0}$, true positives is $C_{1,1}$ and false positives is $C_{0,1}$.

There are two possible predicted classes: "yes" and "no". When we were predicting the presence of a brain tumor, for example, "yes" would mean that it is present, and "no" would mean they don't have the brain tumor.

True positives (TP): These are cases in which we predicted yes (they have the disease), and they do have the disease.

True negatives (TN): We predicted no, and they do not have the disease.

False positives (FP): We predicted yes, but they do not actually have the disease. (Also known as a "Type I error.")

False negatives (FN): We predicted no, but they do have the disease. (Also known as a "Type II error.")

Table 2. Confusion Matrix example

	Predicted : NO	Predicted: YES
Actual : NO	TN = 0 0	FP = 0 1
Actual : YES	FN = 1 0	TP = 1 1

The Accuracy is calculated by the following equation:

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

The Recall is calculated by the following equation:

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

The Precision is calculated by the following equation:

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

1. Results from Machine Learning Algorithms

The diagonal elements represent the number of points for which the predicted label is equal to the true label, while off-diagonal elements are those that are mislabeled by the classifier. The higher the diagonal values of the confusion matrix the better, indicating many correct predictions. The fig.6 shows the confusion matrix without normalization by class support size (number of elements in each class).

This kind of normalization can be interesting in case of class imbalance to have a more visual interpretation of which class is being misclassified. We can use the seaborn package in Python to get a more vivid display of the matrix. Therefore, in our model the following confusion matrix is achieved according to the calculated results of accuracy, precision and recall for Naive Bayes, SVM and Random Forest Classifiers.

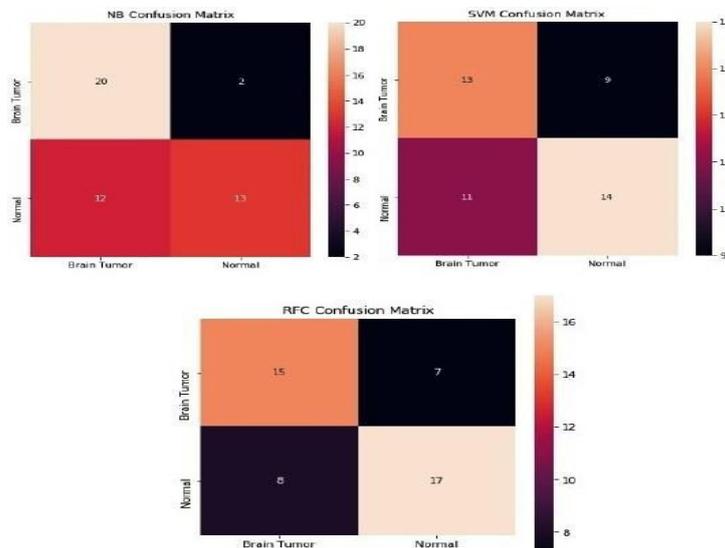


Figure 5:Resulting Confusion Matrices

2. Results from Convolutional Neural Network

We import the data from a .csv file and then split it across three sets: Train, Validation, and Test. The train data will be used to train the model while the validation model will be used to test the fitness of the model. After each run, we can adjust the hyper parameters such as the number of layers in the network, the number of nodes per layer, number of epochs, etc. These adjustments are mostly use on a trial and error basis and the visualization tools, such as the plots given out by Matplotlib, do help in getting to desirable results. The Test Set must not be involved in the training exercise at the parameter or hyper parameter level. In case we intentionally or unintentionally use the test data for training purpose, the test data will not be able to accurately predict the generalization power of the model.

Of the entire data set, 70% is treated as the training set, 10% as the validation set, and 20% as the test set. The training data set is fed to the three-layered Neural networks; with the first two layers having four nodes each and the output layer with just one node. The loss and accuracy data of the model for each epoch is stored in the history object. 1 epoch is an instance where your model has seen the entire training set once. For 15 epoch or 15 times, if there is no increase in validation accuracy, we stop the training.

We need to have optimizer and loss functions for compiling the model. Optimizer is the one that is controlling the loss functions and the steps of the gradient descent i.e. to achieve minimal loss. At the end of 15 epoch we can have a graph that shows how the accuracy has increased or decreased, both validation and training as well as loss. This is how CNN models are trained.

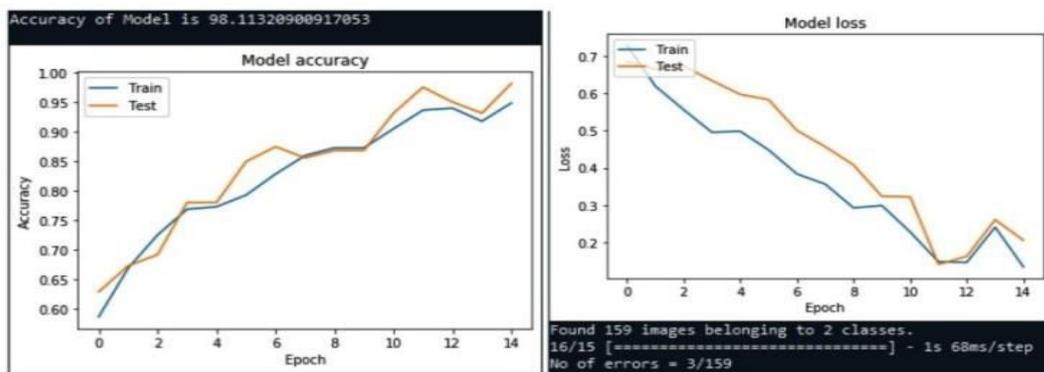


Figure 6: Accuracy and Loss Graph

3. Confusion Matrix

The confusion matrix follows the same principle as that of machine learning algorithms where the values are calculated from the mentioned formulas and by importing seaborn and sklearn the matrix is constructed in a 4x4 form without normalization by considering the validation class and predicted class values.

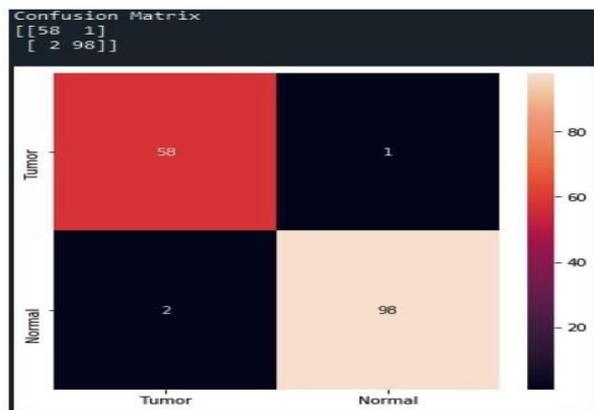


Figure 7: Confusion Matrix obtained

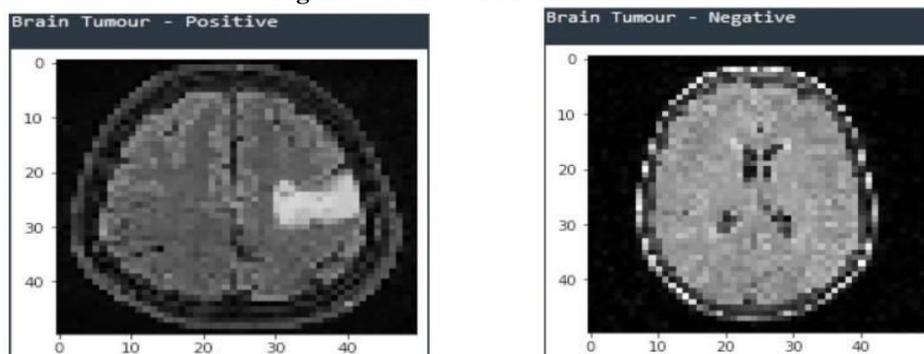


Figure 8: Brain Tumor detection using CNN



4. Result Comparison

Naive Bayes algorithm gets the accuracy of 70.21% with a recall of 0.52 with a precision score of 0.86666. SVM algorithm gives an accuracy of 57.44% with a recall of 0.56 with a precision score of 0.60869. Random Forest algorithm provides an accuracy of 68.08% with a recall of 0.68 with a precision score of 0.70833. Compared to all of these algorithms, the CNN algorithm that is employed by us in this paper provides us with an accuracy of 98.11% with a recall of 0.98 and precision score of 5. The comparison of all these is depicted in Fig.9 as below. We conclude out of Naïve Bayes, Support Vector Machine, Random Forest Classifier and Convolutional Neural Network the classifier that provides the most accurate prediction on whether brain tumor is positive or negative is Convolutional Neural Network.

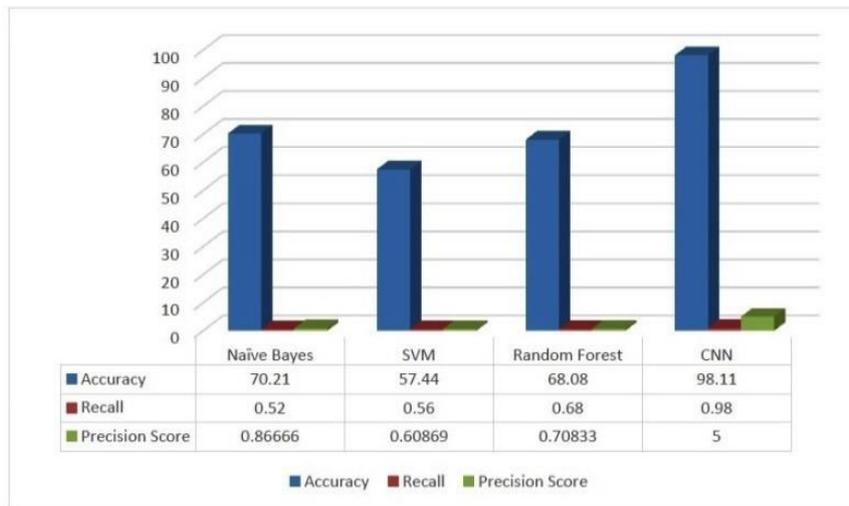


Figure 9: Comparison of all Classifiers

VI. CONCLUSION

The main goal of medical image processing is to identify accurate and meaningful information using algorithms with minimum error possible. Various segmentation methodologies are explored in the paper. It can be concluded that the algorithms and the parameters used in the proposed system are all meant to increase the efficiency of the system by achieving better results.

Of the various classification methods studied, it was experimentally found that the convolution neural networks give the best classification accuracy. Accuracy and reliability are of utmost importance in tumor diagnosis, as a patient’s life depends on the results predicted by the system.

VIII. FUTURE ENHANCEMENT

Encouraged by the results, future work will involve the improvement of classification result and overall accuracy. The number of output classes can also be increased if more data is available. With a more extensive and diverse dataset, the overall classification accuracy can be dramatically increased. Another approach to improve the result would be to increase the number of hidden layers of the neural network. By increasing the number of hidden layers, the weights will be better adjusted and thus increase the classification. One can also do fine tuning and transfer learning approaches to better tune the model based on already trained models.

REFERENCES

- [1] J. Seetha, and S.S. Raja (2018) “Brain Tumor Classification Using Convolutional Neural Networks”, Biomedical and Pharmacology Journal 11(3), 1457-1461.
- [2] B. R.Pushpa, F.Louies (2019) Detection and classification of brain tumor using machine learning approaches”, International Journal of Research in Pharmaceutical Sciences.
- [3] M. Siar, and M.Teshnehlab (2019)Brain Tumor Detection Using Deep Neural Network and Machine Learning Algorithm, 9th International Conference on Computer and Knowledge Engineering (ICCCKE).



- [4] H. E.Abdalla, M. Y.Esmail. (2018) Brain Tumor Detection by using Artificial Neural Network, International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Electronic ISBN:978-1-5386-4123-1.
- [5] E.M. Husein, and D.M.A. Mahmoud (2012)Brain Tumor Detection Using Artificial Neural Networks”, Journal of Engineering and Computer Sciences(ECS), 13.
- [6] Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M. and Larochelle, H (2017) Brain tumor segmentation with Deep Neural Networks, Medical Image Analysis, 35, 18-31
- [7] Li, M., Kuang, L., Xu, S. and Sha, Z. (2019)Brain Tumor Detection Based on Multimodal Information Fusion and Convolutional Neural Network, IEEE Access, 7, 180134-180146.
- [8] Ucuzal, H., Yaşar, Ş, and Çolak, C. (2019)Classification of brain tumor types by deep learning with convolutional neural network on magnetic resonance images using a developed web-based interface, 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2019
- [9] M.Jasmine, P.Satyanarayana (2017) A Relative Study of MRI Based Brain Tumor Segmentation Methods”, International Journal of Pure and Applied Mathematics, 114(10), 133-139.
- [10] K. Rajesh Babu, U. Sai Deepthi, A. SudhaMadhuri, P. Sai Prasad, S. shameem (2019). Comparative Analysis Of Brain Tumor Detection Using Deep Learning Methods, International Journal of Scientific & Technology Research, 8(12),2019.
- [11] T.Hossain, F.S. Shishir, M.Ashraf, M.A. Al Nasim, and F. M. Shah (2019)Brain Tumor Detection Using Convolutional Neural Network, 1st International Conference on Advances in Science, Engineering and Robotics Technology, 1-6.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor

**Impact Factor:
7.512**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details