



Implementation of an Optimized CNN Based Predication Model for Brain Cancer Prediction and Feature Extraction Using ML

Tanusree Saha^{1*}, Tanmoy Das², Dr. Kumar Vishal³, Soumyabrata Saha⁴,

¹JIS College of Engineering, Kalyani, Nadia & Research Scholar, Magadh University, Bodh Gaya.

²JIS College of Engineering, Kalyani, Nadia,

³Department of Mathematics, Magadh University, Bodh Gaya

⁴JIS College of Engineering, Kalyani, Nadia

*Corresponding author:

TanusreeSaha

Research Scholar, Magadh University, Bodh-Gaya & Department of Information Technology,
JIS College of Engineering,
Kalyani 741235, West Bengal, India.

E-mail address: tanusreesah@gmail.com

Tel: + 918617728548;

ABSTRACT:

A tumor is a tissue formed by an accumulation of aberrant cells. These abnormal cells consume the healthy bodily cells, obliterate them, and continue to swell. Brain tumors are one of these tumors. There are two categories of brain tumors: benign and malignant. A cancerous tumor is malignant. Deep convolutional neural networks have mostly developed on these applications in the extremely popular machine learning domain of image categorization. The performance of these networks in terms of prediction accuracy is astounding. In this study, brain magnetic resonance imaging (MRI) pictures were used to create a convolution neural network (CNN) based prediction structure to find tumors. For picture categorization issues, a framework for machine learning that is precise and understandable is given. Three fundamental conditions that in a framework for feature extraction and explanation extraction domain, the ability to assess the eminence of any model's prediction explanation aimed at whichever application has been assimilated.

KEYWORDS: CNN, Image Classification, Feature Extraction, SVM, Logistic Regression.

INTRODUCTION:

In the recent years, a significant number of children and adults have died from brain tumors, one of the most terrible natures of cancer. According to WHO (World Health Organization) guidelines, there are currently 700,000 people living with brain tumors, and 86,000 of those cases have been diagnosed as of 2019. However there have been 16,830 brain tumor-related fatalities since the year 2019 and the average survival rate is 35%. In this study, brain Magnetic Resonance Imaging (MRI) pictures were utilized to create a Convolution Neural Network (CNN) to identify a tumor. Deep convolution neural networks have mostly developed on these applications in the extremely popular machine learning domain of image categorization. These networks are able to perform admirably in terms of forecasting precision, but because they are unable to comprehend their internal workings and provide an explanation for the primary assumptions underlying their predictions, they are regarded as "black box" models.

Cancer has proved itself to be one of the greatest problems in the current world. Cancer is uncontrolled increase in the number of cells through division without any differentiation. Growth of cancerous cells can lead to various dysfunction of the body. Histopathological images have a crucial part in identifying and predicting cancer's prognosis. These photos feature morphological disease progression signs that are both diagnostic and prognostic. Strong computer-assisted analytical methods have been developed over the past ten years thanks to dramatic increases in computing power, advancements in image analysis software, and the introduction of whole slide digital scanners that enable high resolution digitization of tissue histopathology slides. To supplement the pathologist's perspective, algorithms for disease detection, diagnosis, and prognosis prediction have begun to be developed, similar to the function of algorithms for computer-aided diagnosis in

medical imaging. Processing of these histopathological images requires a lot of computational power, so to avoid any chance of missing out on the complex morphological information stored in them, the histopathological images are sliced into tiles to obtain high resolution patches. Maintaining an aided approach to histopathological image is highly prescribed in many papers. With these advances, the domain of research on applying machine learning techniques in Cancer detection and prognosis can be broadly classified into two types: i) Cancer Detection: In this domain vast majority of works are done based on using ML methods to identify, classify, detect or distinguish tumors and cancers; ii) Cancer Prediction: In cancer prediction or prognosis there are two main foci as mentioned in the works of Konstantina Kourou et al.

- Cancer Susceptibility Prediction: Prior to the onset of the disease, research is conducted to determine the risk or possibility of developing a certain type of cancer. a prognostic forecast to support the pathologist's judgment.
- Cancer Recurrence Prediction: The main focus is on employing machine learning techniques to forecast cancer risk following the resolution or diagnosis of a condition e. Cancer Survivability Prediction: Progression after the diagnosis of the disease, the life expectancy, the drug sensitivity all these aspects are predicted in this domain.

In this paper we basically we have around 2500 MRI (Magnetic Resonance Imaging) Images of Brains. We are conducting an experiment. In that Experiment we are trying to predict the Brains which have tumor in it and also the Brain images which does not have tumor in it. We combine three well-known machine learning algorithms—Logistic Regression, Support Vector Machine, and Convolution Neural Network—and determine the accuracy of each one separately.

LITERATURE REVIEW:

A variety of machine learning techniques have been applied in recent years in several research studies aimed at the categorization, the detection and treatment of cancer. Rehman and others created BU-Net to divide and categorize the various locations of brain tumors. For segmenting brain tumors, It was suggested to use their concept with updated encoder-decoder architecture. 2017 and 2018 BraTS, the suggested BU-Net architecture was tested. Compared to alternative segmentation models and the U-Net baseline architecture already in use, BU-Net has shown to be a substantial improvement. Radiology Because MRI imaging sub-regions can change tumour localization, researchers proposed a deep learning method that takes this into account. Subsequently, they used a typical 3D convolutional neural network to classify the tumor segments into subtypes (CNN). Popular metrics including the mean square error, the Hausdorff distance at percentile 95 (HD95), and the dice score coefficient were employed to assess performance. The results indicate that the proposed method can reliably segregate tumors and forecast survival rates. In order to identify breast cancer, Bashir et al. combined five classifiers: Support vector machines, naive Bayes, decision trees using the Gini index, decision trees using information gain, and memory-based learner. The final forecast was then made using the ensemble technique based on weighted votes. To increase the prediction accuracy, a Four datasets pertaining to breast cancer were also subjected to a variety of preprocessing and feature selection strategies. With averages of 85.23%, 86.18%, and 76.68% for accuracy, precision, and recall, the suggested ensemble classifiers produced outstanding results. Few datasets were utilized to assess the model's performance, nevertheless. Computational instability may result from using the a large dataset's proposed ensemble with numerous features. Kumar et al. offered the ANOVA, Kruskal-Wallis, and Friedman tests as illustrations of statistical procedures (tests) based on MapReduce to pick pertinent features in order to address the dimensionality issue. After feature selection, the Using a proximal support vector machine (mrPSVM) classifier built on the MapReduce framework, the microarray data. These methods were put into practice using the Hadoop platform. Using microarray datasets of different dimensions, a comparison of these feature selection methods was done. According to the experimental findings, an ensemble of the mrPSVM classifier and several feature selection techniques created models with higher accuracy than others. As a result, the suggested model handled huge data successfully, however it could only understand biological microarray data. Lastly, Jain et al. suggested a two-phase hybrid model (iBPSO) for cancer classification that included correlation-based features selection (CFS) and improved-binary particle swarm optimisation. The proposed model selects a low-dimensional set of prognostic genes from biological samples of binary and multi-class tumours using the naive Bayes classifier. Using eleven benchmark microarray datasets, the model was examined and tested. The results of the tests revealed that, in terms of classification accuracy and the number of chosen genes, the model performed better than numerous well-known methodologies. Pradana et al. provide a method employing binary particle swarm optimisation (BPSO) as a feature selection and C4.5 decision tree

as a classifier to investigate cancer diagnosis based on microarray data. K-Means is used to discretize the data for the decision tree rule model. The model was successfully able to identify the most crucial elements and improve accuracy when BPSO and decision trees were applied. For C4.5 and BPSO, the model's accuracy was 54% and 99%, respectively. In order to find highly significant genes in microarray gene expression datasets, Shukla et al. also suggested a novel filter-based gene selection approach.

For the diffuse large B cell lymphoma (DLBCL) dataset, the suggested method was assessed using well-known classification approaches such as support vector machine, naive Bayes, k-nearest neighbour, and decision tree. The proposed model could compete successfully, according to experiments, and made outstanding predictions for accuracy, precision, sensitivity, measure, and ROC value.

A unique cuckoo search using a crossover algorithm was created by Sampathkumar et al. to successfully categorise various cancer subtypes. In tests using benchmark cancer gene expression, the model surpassed CS and other well-known approaches, according to the results. The relief algorithm was applied by Kilicarslan et al. for feature ranking and dimension reduction. Support vector machines (SVM) and convolution neural networks (CNN) were then employed for prediction using the most crucial features. The experimental findings demonstrate that the suggested strategy may raise the precision of SVM and CNN classification algorithms. Last but not least, Lee et al. proposed a novel using a multivariate feature ranking technique, gene selection in microarray data can be more effective. By incorporating the formal concept of relevance into a Markov blanket, the suggested method generated a novel feature ranking mechanism (MB). The outcomes demonstrate that the model is effective in classifying high-dimensional microarray data. All of the techniques in the above list address the two main problems of brain cancer classifiers and the curse of dimensionality, which have been discussed thus far. However, here in this paper we are trying to build different Machine Learning and Deep learning algorithms and try to find their accuracy, value loss function and value loss function also we are trying to compare them and reach to a specific conclusion that which is the best algorithms to work with the medical images data.

METHODOLOGY:

In this paper we are using three techniques to predict brain cancer. We can take MRI (Magnetic Resonance Imaging) Images of the patient's brain. Then practice deep learning and machine learning algorithms to detect any patches or tumors in the brain. Then upon the condition of the brain we can predict if it is going to be a brain cancer or not. Second there is technique in which using medical background of the patient we can predict if the patient is going to have brain cancer in near future. Third way and the most effective way to predict the Brain Cancer to study "Genomic Sequence" of the patient and using Deep Learning check whether the combination of the genes can have Brain Cancer in future.

In this study, we are doing a comparison between three well known machine learning and deep learning algorithms that are:

- I. Logistic Regression (Machine Learning),
- II. Support Vector Machine (Machine Learning)
- III. Convolution Neural Network (Deep Learning).

Logistic Regression: Logistic regression is a statistical method used to analyze and model binary or categorical outcomes. It is a type of regression analysis that is used when the response variable is binary, meaning it can take only two possible values, such as 0 or 1. The goal of logistic regression is to model the probability of the binary response variable as a function of one or more predictor variables.

In logistic regression, the response variable is modeled using a logistic function, which is an S-shaped curve that maps any real-valued input to a value between 0 and 1.

Support Vector Machine: Support Vector Machine (SVM) is a powerful machine learning algorithm used for classification and regression analysis. It is widely used in various fields such as finance, healthcare, bioinformatics, and text classification.

In SVM, the goal is to find a hyperplane that maximally separates the data points of different classes in a high-dimensional space. The hyperplane is selected in such a way that the distance between it and the nearest data point of each class is maximized, which is called the margin. If the data is not linearly separable, the SVM uses a technique called kernel trick to map the data into a higher-dimensional space, where it may become linearly separable. The most common types of kernels used are linear, polynomial, and radial basis function (RBF) kernels.

The decision boundary of SVM is determined by support vectors, which are the data points closest to the

hyperplane. These support vectors are used to calculate the margin and the decision boundary. SVM has several advantages over other classification algorithms. It works well with high-dimensional data, can handle both linearly and nonlinearly separable data, and has a strong theoretical foundation.

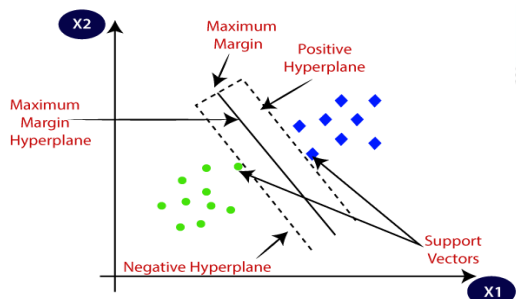


Fig: SVM Hyper plane

Convolution Neural Network: Convolution Neural Network, or CNN for short, is essentially a deep learning algorithm. To learn new things, it duplicates animal brain cells, or neurons. If we provide CNN a lot of data, it will essentially try to find a pattern between them and categorize them as such. CNN mostly submitted applications for computer vision. Here, we use the convolution layer to extract a certain feature from the image.



Fig.5: Lionel Messi

Here we have image of Lionel Messi. If we extract face feature and pass it through convolution layer. It will basically check the feature of the face in every pixel value. As compare to ANN (Artificial Neural Network), the feature will be checked at the exact pixel value where it finds the matched feature.

RESULT & DISCUSSION:

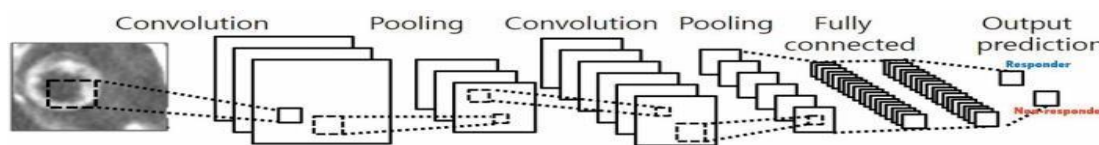
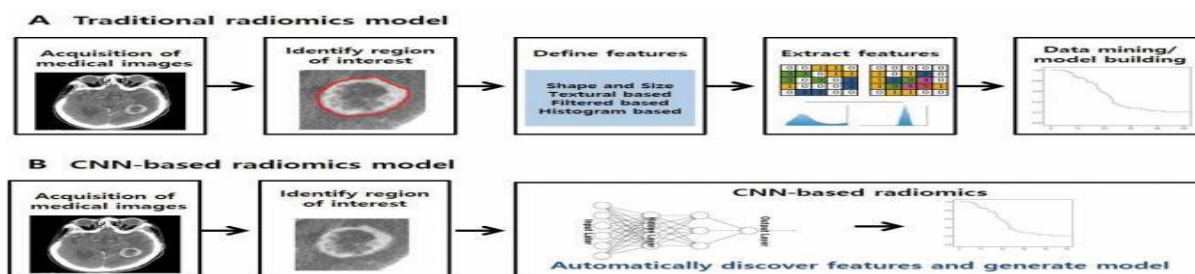


Fig.6: CNN Layers

If we consider CNN there are various components within it. At first when images are feed to the network it goes to the convolution filter. It is going to be responsible for extracting features from the image. Suppose we have an image of a cat. Cat has eyes, nose, ears etc. these features will be extracted by this layers and updated into a matrix form.

Then we have pooling layers. As we have passed our image through multiple convolution filter, the dimension matrix will increase. So to reduce dimension we use pooling layer. We take 2 X 2 submatrix which is industry standard then we take the max value or average value of the matrix and create a pooling matrix. In our experiment we have used Maxpool function. The next method we have used is padding layer. So, if we pass an image from convolution filter that will create a matrix which is smaller than the actual image but we have some feature that can present at the corner of

the image. To solve such problem we pad 0's and create same size matrix as the original one. Now we can extract features from every corner of the image. So as we can see we have got the matrix representation of an image but we can't pass it through a perception because it only allows 1 X N size matrix. Suppose we have 20 X 20 matrix this layer will flatten it and convert to 1 X (20X20) that is 1 X 400 size matrix. After these steps the mechanism is ready to be fully connected and give outputs.



Optimization of CNN: After building this model a big task was to get the best accuracy possible for the algorithm. For this challenge we incorporate a unique way. In CNN we have to feed our trained data many times to the model. This reiteration called as epochs. One epoch means one iteration. Now for best accuracy we have chosen a random number that is 32 and calculate the accuracy and note it. After that we have increase the number of epochs by a constant number in our case that is 8. Now check the accuracy and compare with the previous one if accuracy of the previous one is lesser than update the previous accuracy with new one. Do this until accuracy decreases. When the accuracy decreases divide the constant by 2 and update the constant. In our case we divide 8 by 2 and 4 as output. Now we reduce the epoch's value by 4 and try to calculate the accuracy. Do these steps until find the optimal epoch size where the accuracy is highest. We have conducted our experiment on three different algorithms.

- Logistic Regression
- Support Vector Machine
- Convolution Neural Network

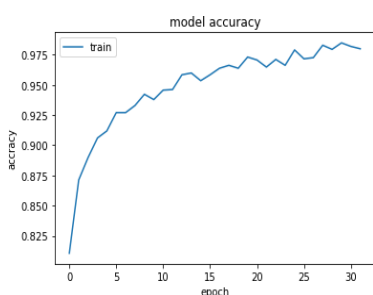
Table 1: Accuracy Analysis

	Logistic Regression	Support Vector Machine	Convolution Neural Network
Accuracy	0.9070	0.9290	0.9820 (epochs 32)

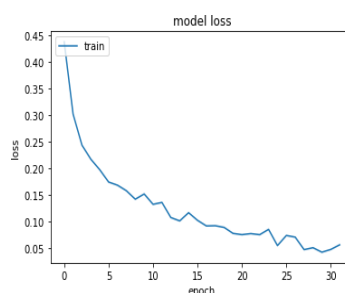
DETAIL ANALYSIS FOR CNN:

1. Calculation for 32 epochs:

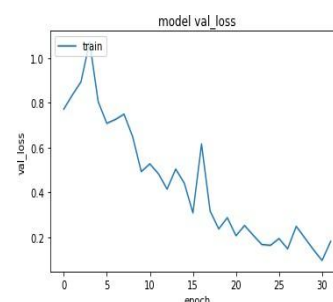
Accuracy: 0.9820



Loss: 0.0522



Value_Loss: 0.1812

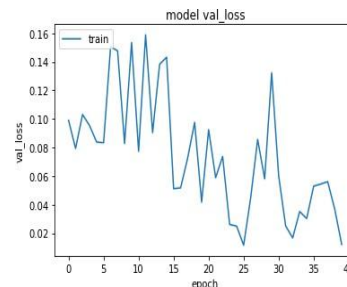
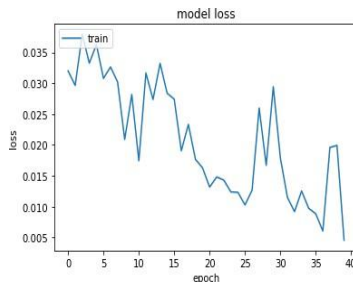
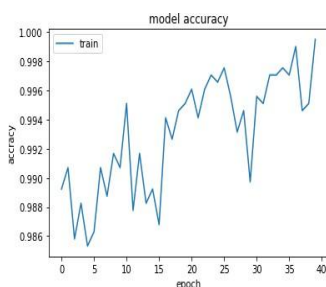


2 .Calculationfor40epochs:

Accuracy:0.9995

Loss:0.0046

Value loss:0.0121

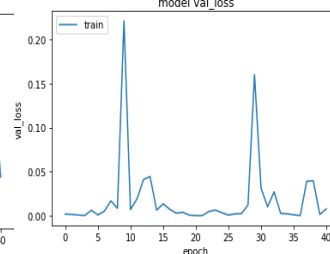
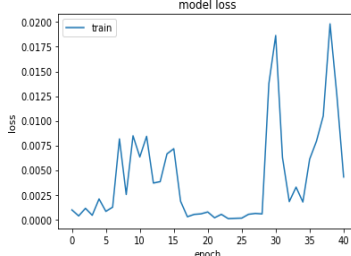
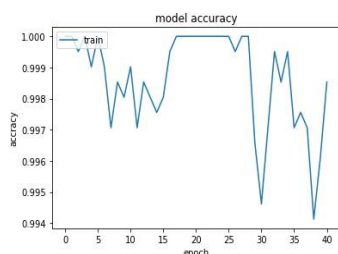


3. Calculationfor41epochs:

Accuracy:0.9985

Loss:0.0043

Value_loss:0.0079

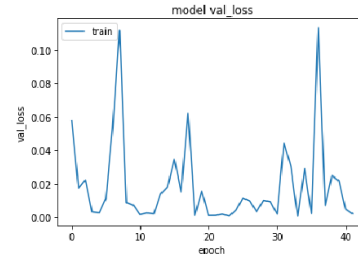
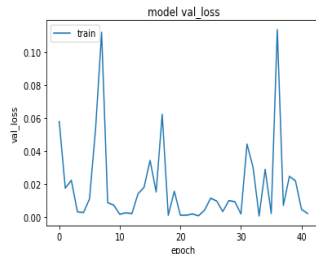
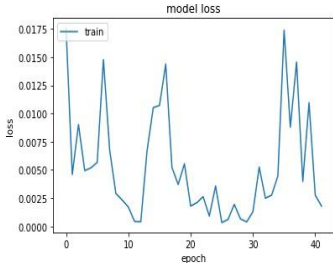


4. Calculationforepochs42:

Accuracy:0.9990

Loss:0.0018

Value_loss:0.0020

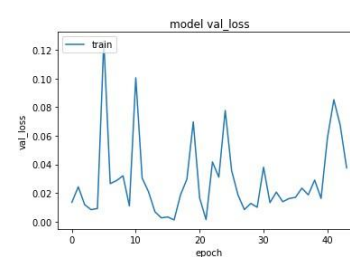
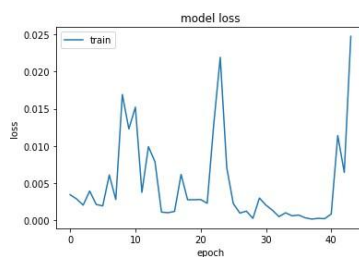
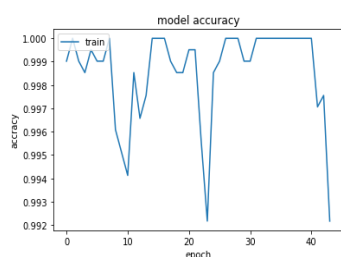


5. Calculationfor44epochs:

Accuracy:0.9922

Loss: 0.0248

Value_loss:0.0376

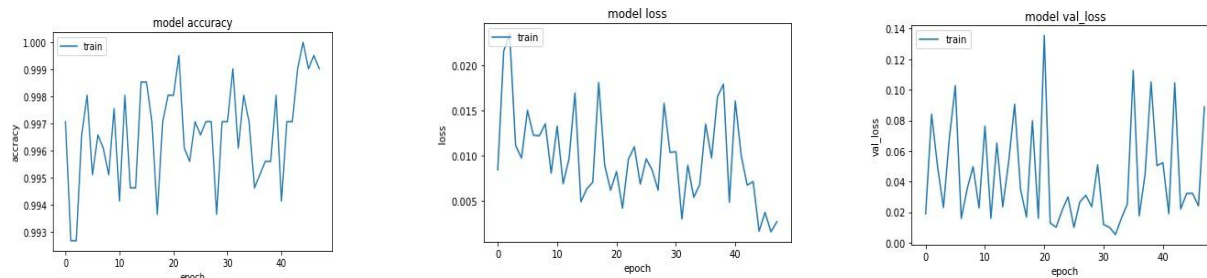


3 .Calculationfor48epochs:

Accuracy:0.9990

Loss: 0.0027

Value_loss:0.0888



After calculating these six different numbers of epochs we can clearly see that the highest accuracy we get is 0.9995. This accuracy we got for calculating 40 epochs.

CONCLUSION:

In this work, a framework for precise, reliable, and explicable prediction of picture classification tasks was provided, offering three different forms of explanation outputs with regard to the audience domain.

Unlike the majority of techniques, our method is intrinsically interpretable and provides clear justifications based on inputs from high-level feature representations that are retrieved from images. These characteristics are meant to characterize the characteristics that each pixel in a picture possesses, such as the degree of texture irregularity, the size and shape of an object, etc. A vital limit of our method is that some of these components may only make sense to certain human experts and image analysts. Nonetheless, we tried to plain-spoken human language; qualitatively explain what these features mean to increase the appeal and viability of the explanation output from our algorithm.

Last but not least, even the least significant features were used in our trials. We made an effort to scale back the quantity of features. In this dataset by looking at the associations among the characteristics as well as their relevance and employing a variety of feature selection algorithms. No feature was eliminated since doing so would have negatively impacted the performance of the prediction model as a whole. Our main objective was to create and disseminate an intelligible CNN-based Machine Learning an image classification framework; therefore we want to be clear that processing feature selection was outside the scope of our work.

CONFLICT OF INTEREST:

The authors have no conflicts of interest regarding this investigation.

REFERENCES:

- Emmanuel Pintelas, Meletis Liaskos, Ioannis E. Livieris, Sotiris Kotsiantis, Panagiotis Pintelas, Explainable Machine Learning Framework for Image Classification Problems: Case Study on Glioma Cancer Prediction, *J. Imaging* 2020, 6(6), 37; <https://doi.org/10.3390/jimaging6060037>, 28th May, 2020.
- Rawat, W.; Wang, Z. Deep convolution neural networks for image classification: A comprehensive review. *Neural Comput.* 2017, 29, 2352–2449.
- Park JG, Lee C (2009) Skull stripping based on region growing for magnetic resonance brain images. *Neuroimage* 47:1394–1407.
- Yasmin M, Mohsin S, Sharif M, Raza M, Masood S (2012) Brain image analysis: a survey. *World Appl Sci J* 19:1484–1494.
- Somasundaram K, Kalaiselvi T (2010) fully automatic brain extraction algorithm for axial T2-weighted magnetic resonance images. *Comput Biol Med* 40:811–822.
- Sharif M, Amin J, Nisar MW, Anjum MA, Muhammad N, Shad SA (2020) A unified patch based method for brain tumor detection using features fusion. *Cogn Syst Res* 59:273–286.
- Spyridon Bakas, Mauricio Reyes, Andrés Jakab, Andreas Mang, Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge, 2019
- Subhashis Banerjee, Somosmita Mitra, Yoichi Hayashi, Volumetric brain tumour detection from MRI using visual saliency, DOI:10.1371/journal.pone.0187209, 2017
- Joeky T Senders 1, Patrick C Staples 2, Aditya V Karhade 3, Mark M Zaki 3, William B Gormley 3, Marike LD Broekman 1, Timothy R Smith 3, Omar Amaout, Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review, 2018
- Dokyoon Kim 1, Hyunjung Shin, Young Soo Song, Ju Han Kim, Synergistic effect of different levels of genomic data for cancer clinical outcome prediction., 2019
- Patric Fulop, A. Manataki, Predicting Survival After Surgery For Brain Tumour Patients: A Machine Learning Study On Clinical Data And Molecular Data, Corpus ID: 210905861, 2019.
- D. Lu, Q. WENG, A survey of image classification methods and techniques for improving classification performance, *International Journal of Remote Sensing*, 2007, Vol. 28, No. 5, pp. 823–870.
- Le Hoang Thai, Tran Son Ha, Nguyen Thanh Thuy, Image Classification using Support Vector Machine and Artificial Neural Network, *I.J. Information Technology and Computer Science*, 2012, 5, 32–38. Published Online May 2012 in MECS (<http://www.mecs-press.org/>) DOI:10.5815/ijitcs.2012.05.05
- Sai Yeshwanth Chaganti, Ipseeta Nanda, Koteswara Rao Pandi, Tavva GN RSN Prudhivith, Niraj Kumar, Image Classification using SVM and CNN, Conference Paper - March 2021, DOI:10.1109/ICCSEA49143.2020.9132851.