ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

Brain Tumor Prediction using Data Mining

Yogita Patil[1], Revati Pimparkar[2], Sukanya Sonar[3], Nidhi Zope[4]

U.G. Student, Department of Computer Engineering, SSBT's COET Bambhori, Jalgaon, India[1]

ABSTRACT: In the recent era there has been a rapid growth in the availability of medical databases and medical imagery in the past few years, and the uncertainty involved in the effective prediction of diseases from these databases made the research community to take up the challenges in this domain**.** The Central Nervous System of mankind being is mainly composed of Spinal Cord, Brain and Neurons. Largest part of brain is Cerebrum. The human brain throws good number of challenges to the research community. The Brain Tumor or the Intra Cranial Neoplasm is formed due to the irregular growth of cells in the brain. This sort of irregular growth of cells in brain damages frontal, temporal and parietal lobes thereby results in abnormal behavior. Machine learning algorithms provide computers with the ability to learn without being explicitly programmed. The clinical brain data set was analyzed effectively using machine learning algorithms and made conclusions on the results. In this paper we applied Hybrid Data Mining methods which in turn consists of Clustering, Classification and Association techniques and further analyzed the results using some statistical techniques.

I. INTRODUCTION

Human brain represents only 2% mass of total body but uses 20% body's energy [1]. Brain controls all the activities of the human body. So the brain needs to operate with its maximum efficiency. Now-a-days, a lot of people are suffering from brain tumor which causes even death, if not treated at time. Brain Tumor is a cluster of abnormal cells growing rapidly in the brain and clustering is also used for grouping of similar cells [16]. It may occur to any person at any age and appear at any location in the brain. Tumor is further categorized in two: malign and benignant. Benignant tumors have homogeneous structure and don't contain cancer cells while malign have heterogeneous structure and contain cancer cells. Benign tumors are either radio-logically or surgically destroyed and have rare chances of grow back. Malignant are life threatening tumor and can be treated by chemotherapy, radiotherapy or their combination. So, need to diagnose the tumor at an early stage is essential for future treatments.

MRI (Magnetic Resonance Imaging) has proven out as a powerful tool in detection of brain tumor with the help of MR Images. It is a non-invasive technique which produces very detailed 2D and 3D images of the organ inside the brain in every direction. As the large amount of data provided through MRI technique, so it is impractical to develop a method which can classify the images in normal or abnormal through human inspection. [2]

Data Mining has been known for evolving out some important features from large amount of data. Due to this specialization of data mining, this field is used in combination with medical science for the accurate diagnosis of the patient disease. A no. of classification methods has been evolved under data mining. In order to achieve best accuracy model, we will compare the accuracy determined by the different classification models given as: SVM, decision tree classifier, Naïve Bayesian Classifier and KNN algorithm on the specific datasets: primary tumor dataset obtained from the UCI Web Repository.

II. LITERATURE SURVEY

V. P. GladisPushpaRathi et al. [1] stated that "feature extraction is a technique of catching image visual content in raw form and its storage in reduced form that facilitates decision making process, this type can be treated as pattern classification. This technique clubs features such as shape, texture, intensity and classifies the tumor as, gray matter, white matter, CSF, normal and abnormal area. One of the classification techniques SVM serves as measure for comparison between linear verses non-linear techniques. To reduce the number of features selected LDA and PCA techniques are used". Roopali R. Laddha et al. [2] stated that "Exact detection of location and size of brain tumor plays a vital role in diagnosing it. An effective algorithm was proposed for brain tumor discovery which is based on morphological and

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

segmentation operators. Quality of the image derived should be enhanced and it is subjected to morphological operators so as to detect the tumor. By using MRI and CT images redundant and complementary information is captured by implementing wavelet based technique". Varsha Kshirsagar et al. [3] stated that "Simple Algorithms are used for detection of shape and range of tumor through brain images further computer aided methods like segmentation (detection) can be implemented to detect. Accuracy of tumor detection enhances on using segmentation techniques. Time for analyzing all such images also reduces. This entire process outcome would be detection of brain tumor with input being MRI along with its size and position. From the total area of the tumor derived from a specified cluster stage and the intensity of the tumor is defined". Punithavathy Mohan et al. [4] stated that "This deals with image segmentation, its enhancement, segment extraction and the classification of the Brain MRI. Proposed algorithm is CLAHE, k-means and ACO (Ant Colony Optimization). ACO is utilized for segmenting the image and k-means algorithm is utilized for classifying tissues as abnormal and normal from the brain MRI with very less time complexity and good level of accuracy". Jay Patel et al.[5]stated that "various clustering techniques like Fuzzy C-means, k-means Clustering have been used for different segmentation methods namely Thresholding, Region growing and Mean shift in MRI are reviewed".

Alan Jose et al. [6] stated that "Using fuzzy c-means, k-means clustering segmentation carried out results in detection of brain tumor and its position. Fuzzy c-means performance is much better when compared with other techniques for segmentation. This process also enables in understanding the tumor stage, its severity and whether it is at a curable state or not". Shubhangi S. Veer et al. [7] stated that "The global threshold and watershed segmentation are useful techniques for the segmentation of MRI or CT scan images. The global threshold technique separates out diseased region by considering a single threshold point and further binary image is derived from the gray scale image. Watershed segmentation technique also separates out tumor region successfully from the brain tumor MRI region. From these segmented images it is possible to get the detail information about the tumor location, its shape and area of the extracted tumor region is measured in terms of the number of white pixels in the segmented image".

Roshan G. Selkar et al. [8] stated that "Brain tumor identification helps in the discovery of exact shape, size, location and boundary extraction of the tumor. The process has a total of three stages to identify and then to segment the brain tumor. Firstly the quality of image scanned has to be improved later the morphological operators were applied to identify the tumor in the image scanned. After that edge detection operator is applied for boundary extraction and to find the size of the tumor". Gauri P. Anand gaonkar et al[9] stated that "In order to locate tumor from the MR images derived different segmentation methods can be used. A method has been proposed to carry out the same utilizing an algorithm called Fuzzy C-Means and then to calculate the A.K. Mohantly.et.al [11][14] have applied various techniques for the identification of cancerous tissues .Image mining is one of the technique through which we can mine the hidden data although the image is not in clear. By applying the image mining technique in testing point, space partition is used to categorize the image .S.W. Purnami.et.al
[12] applied data mining technique" Multiple Knot Spline Smooth Support Vector Machine", for identifying the medical diagnosis issues majorly diabetes diseases and heart diseases. Shweta Kharya [13] has discussed about different data mining techniques which are used for the identification of breast cancer classification. N.H. Rajini.et.al [15][18] have projected majorly on two phases which are one of the feature extraction and the other is classification .In the first phase they have obtained the feature extraction which belongs to magnetic resonance imaging using discrete wavelet transformation and in the second phase they have used classification algorithms of feed forward back propagation artificial neural network and k nearest neighbor for automated diagnosis process

B.G. Prasad. et.al[16] have stated various data mining classifiers for the purpose of medical image classification where they have applied J48 decision tree and random forest for grouping the CT scan brain images. The projected developed system is mainly based on texture information of images. V.S. Tseng. et.al[17] projected image categorization methods by applying association rules on the image objects. The developed method is divided into phases. In the first phase construction of object hierarchy and in the second phase applying the multilevel mining algorithms for identifying the image classification rules. A. H. Gondal.et.al evaluated various techniques which highlights new outlook for brain tumor detection from magnetic resonance images.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

III. BRAIN TUMOR DETECTION MODELS

Under this section we will discuss following data mining classification models to detect brain tumor:

A. Decision Tree

Decision trees are the powerful and greedy classification algorithms. The most popular are Quinlan's ID3, C4.5 and CART algorithm. As the name implies, a tree is constructed in a top-down recursive divide and conquer manner. At start, all the observations are at the root. Then the test attributes are selected on the basis of some heuristic or statistical measure, (e.g. information gain). It splits the input observations into two or more subgroups. This process is repeated recursively until the complete tree is constructed. Our main objective is to find the variable-threshold pair which best splits the observations into subgroups. The most commonly used mathematical algorithm for splitting includes Entropy based information gain (used inID3, C4.5, C5), Gini index (used in CART), and Chi-squared test (used in CHAID).

B. Artificial Neural Networks

Artificial Neural Networks are the biologically inspired networks which have the tendency to model extremely complex non-linear functions. ANNs are the highly sophisticated analytical techniques having the capability of learning the existing data. Multi-layer perceptron (MLP) with back-propagation is a supervised learning algorithm which is the one of popular ANN architecture. The use of this algorithm is started by psychologists and neurobiologists in order to develop test computational analogues of neurons. A neural network has a set of connected input/output units where each connection has a weight associated with it. The main usage of neurons in input layer Xi to divide the input signals among neurons in hidden layer. Every neuron j in hidden layer sums its input signals with connections Wij from the input layer and output function given as in following figure

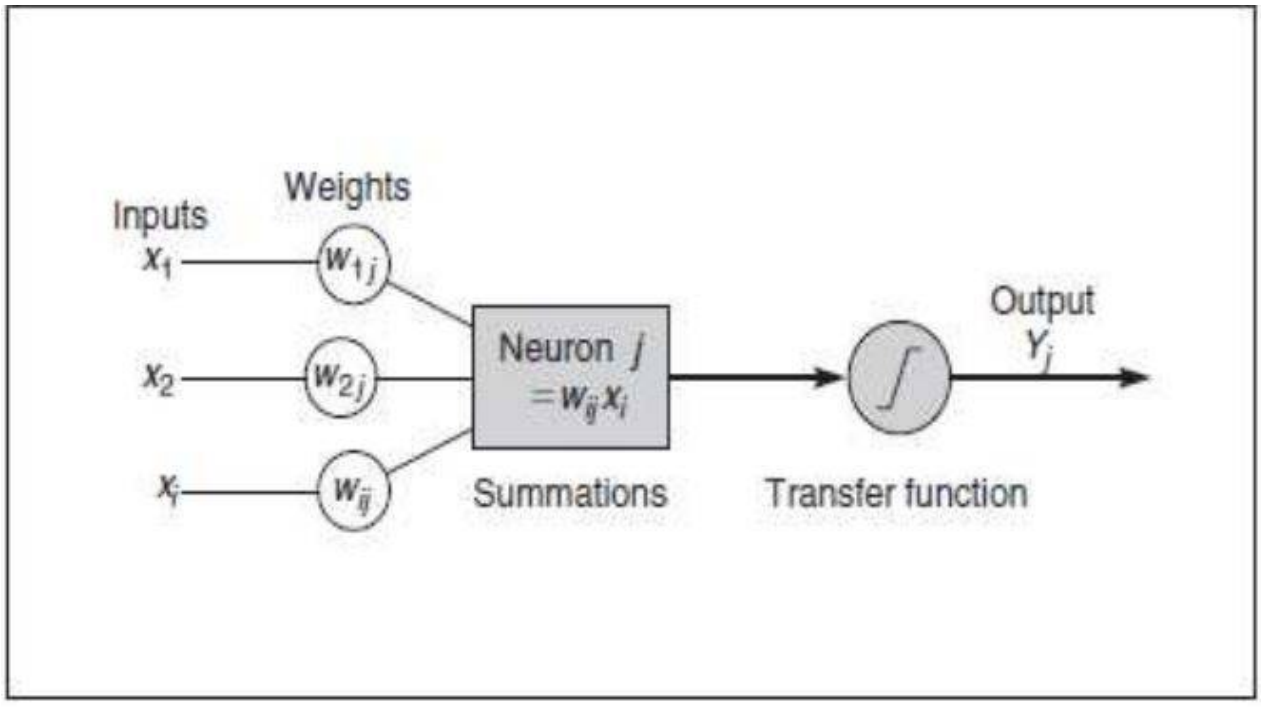

Fig. Artificial Neural Network

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

C. Support Vector Machine

SVM is an up to the minute classification algorithm used for the classification of both linear and non-linear data. This classifier is derived from statistical learning theory given by Vipnik in 1992. SVM classifier approaches the problem by finding out the hyper-plane with largest margin, i.e. maximal marginal hyper-plane. For the data which is not linearly separable, it transforms the original training data into a higher dimension by doing non-linear mapping. By transforming it into high dimensional space, it searches for linear optimal separating hyper-plane. This transformation technique into high dimension always helps in searching for an optimal hyper-plane using support vectors and margins [13]. SVM performs classification by finding optimal MMH and minimizing the classification errors.

D. Naïve Bayesian Classifier

Bayesian classifier demonstrated as a statistical classifier which performs probabilistic prediction, i.e. class membership probabilities. Its foundation based on bayes theorem which described as below given training data X, and posterior probability of hypothesis H is P (H|X):

$$P(H|X)= P(X|H).P(H)/P(X)$$

A Bayesian classifier has close performance with decision tree and ANN classifier. Each training example can affect the probability that a hypothesis is correct either increase or decrease — some prior knowledge can be combined with observed data. Let G be set of training tuples attached with class labels. Each tuple is represented by attribute vector given as $\mathbf{X} = (x_1, x_2,\ldots,x_{i,\ldots,} x_n)$. Let there are a total of *z* classes $C_1, C_2,\ldots, C_z$. Classification is to get the maximum posteriori probability, i.e., $P(C_i|\mathbf{X})$. This can be obtained with the help of Bayes theorem. As P(X) is constant for all classes, only $P(C_i \mid \mathbf{X}) = P(\mathbf{X} \mid C_i)P(C_i)$ needs to be maximized.

IV. RESULTS

Initially dataset had 17 attributes and 339 records for Primary Tumor data set. Algorithm for attribute selection was applied on dataset to preprocess the dataset. After attribute selection missing values records were identified and were deleted from dataset. After deleting records with missing values we were left with modified records. On these records data mining classification techniques Naïve Bayesian, Decision Tree, Artificial Neural Networks (ANNs) and Support Vector Machine (SVM) were applied. Sensitivity, specificity and accuracy are obtained from the confusion matrix. Confusion matrix is the representation of classification results in the form of matrix.

The goal is to have high accuracy, as well as high precision and recall metrics. These can be easily converted to true-positive (TP) and False-Positive (FP) metrics. Precision means the percentage of your results which are relevant.

Precision= TP/(TP+FP)

On the other hand, Recall refers to the percentage of total relevant results correctly classified by algorithm.

Recall=TP/(TP+FN)

- TP (True Positive): Total Percentage of members classified as class A belongs to class A.
- FP (False Positive): Total Percentage of members of class A but does not belongs to class A.
- FN (False Negative): Total Percentage of members of class A incorrectly classified as not belongs to class A.
- TN (True Negative): Total Percentage of members which do not belong to class A are classified not a part of class A. It can also be given as (100%-FP).

V. CONCLUSION AND FUTURE SCOPE

There are different data mining techniques that can be used for the detection and prevention of brain tumor disease among patients. In this paper four classification techniques in data mining to predict brain tumor disease in patients are

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

compared: Naïve Bayesian, decision tree, Artificial Neural Networks and Support Vector Machine. These techniques are compared on behalf of True Positive Rate, False Positive Rate, Sensitivity, Specificity, Accuracy and Error Rate. In future we intend to improve performance of these basic classification techniques by using some hybrid approach in terms of accuracy and other measuring criteria.

REFERENCES

[1] V.P.Gladis, PushpaRathi, Dr.S.Palani,"Brain Tumor MRI Image Classification with Feature Selection and Extraction Using Linear Discriminant Analysis".
[2] Roopali R.Laddha, S.A.Ladhake, "A Review on Brain Tumor Detection Using Segmentation And Threshold Operations", International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, pg.607-611.
[3] Varsha Kshirsagar, Prof. Jagruti Pancha," Segmentation of Brain Tumour and Its Area Calculation", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 5, May 2014.
[4] Punithavathy Mohan, Vallikannu. AL, B. R. ShyamalaDevi, B.C.Kavitha, "Intelligent Based Brain Tumor Detection Using ACO", International Journal of Innovative Research in Computer and Communication Engineering. Vol. 1, Issue 9, November 2013
[5] Jay Patel and Kaushal Doshi, "A Study of Segmentation Methods for Detection of Tumor in Brain MRI", Advance in Electronic and Electric Engineering, Research India Publications , Volume4, Number 3 2014,pp. 279-284
[6] Alan Jose, S.Ravi, M.Sambath, "Brain Tumor Segmentation Using K-Means Clustering And Fuzzy C-Means Algorithms And Its Area Calculation", International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 3, March 2014.
[7] Shubhangi S. Veer (Handore), Dr. P.M. Patil,"An efficient method for Segmentation and Detection of Brain Tumor in MRI images", International Research Journal of Engineering and Technology, Volume: 02 Issue: 09 , Dec-2015.
[8] Roshan G. Selkar, Prof. M. N. Thakare, "Brain Tumor Detection and Segmentation by using Thresholding and Watershed Algorithm", IJAICT, Volume 1, Issue 3, July 2014.
[9] Gauri P. Anandgaonkar, Ganesh. S. Sable, "Detection and Identification of Brain Tumor in Brain MR Images Using Fuzzy C-Means Segmentation", International Journal of
Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 10, October 2013
[10] Rahil Malhotra, MadhuBahl, Harsimran Kaur, "Brain Tumour Detection for Digital Images", International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 3, Issue 9, September 2013.
[11] A.K. Mohantly, S.K. Lenka, "Efficient Image Mining
Technique for Classification of Mammograms to Detect Breast Cancer," International journal of computer science and communication technologies, vol. 2, no.3, pp.99-106, 2010.
[12]S.W. Purnami, J.M. Zain, & A. Embong, "Data Mining Technique for Medical Diagnosis Using a New Smooth Support Vector Machine," Communications in Computer and Information Science, 2010, pp.15-27.
[13] S. Kharya, "Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease," International Journal of Computer Science, Engineering and Information Technology, vol.2, no.2, pp.55-66, 2012.
P. Rajendran, M. Madheswaran, K. Naganandhini, "An Improved Pre-Processing Technique with Image Mining Approach for the Medical Image Classification," Second International conference on Computing, Communication and Networking Technologies, 2010, pp.183-187
[15] N.H. Rajini, R. Bhavani, "Classification of MRI Brain Images using k- Nearest Neighbor and Artificial Neural Network", International Conference on Recent Trends in Information Technology, 2011, pp.863-868.
[16] B.G. Prasad, A.N. Krishna, "Classification of medical image using data-mining techniques," Advances in Communication, Network, and Computing, 2012, pp.54- 59
[17] V.S. Tseng, M.H. Wang, J.H. Su, "A New Method for Image Classification by Using Multilevel Association Rules," Data Engineering Workshops, 2005, pp.1180- 1187
[18] Y. Zhang. and Z. Dong. "A hybrid method for MRI brain image classification", Expert System with applications, pp. 10049-10053, 2011
[19] A. H. Gondal, and M. N. A. Khan, "A review of fully automated techniques for brain tumor detection from MR images," International Journal of Modern Education and Computer Science (IJMECS), vol. 5, no. 2, pp. 55, 2013