

Challenges Inherent in Building an Intelligent Paradigm for Tumor Detection Using Machine Learning Algorithms



A. S. Shinde, V. V. Desai and M. N. Chavan

Abstract Machine learning is at the heart of the big data rebellion sweeping the world today. It is the science of getting the computers to learn without being explicitly programmed as most of the technological systems are in an insurrection to be operated by intelligent machines capable to make the human like verdict to automatically solve human task with perfect results. Artificial intelligence is the heart of every major technological system in the world today. This paper presents the challenges faced to develop a model to acquiesce excellent results and the different techniques of machine learning; here, we also present the broad view of the current techniques used for detection of Brain tumor in computer-aided diagnosis and an innovative method for detection of Brain tumor by artificial intelligence using the algorithm of k-nearest neighbor which is established on the training a model with different values of k and the appropriate distance metrics are used for the distance calculation between pixels.

Keywords Machine learning • Artificial intelligence • Supervised learning Classification • Regression • Unsupervised learning • K-nearest neighbor

1 Introduction

Brain is the most respective part of the human body. The Brain and the spinal cord comprise of the body's nervous system. The Brain controls the action of the body through this nervous system and this involves receiving information from the sensory organs of the body, interprets the information, and then guides the body's response. Any infection, damage, abnormality to the Brain can cause threat to the normal functioning of the Brain and even human life. Advances in computer-aided

A. S. Shinde (✉) · V. V. Desai
GIT Belgaum, Belgaum, Karnataka, India
e-mail: ashsshinde@gmail.com

M. N. Chavan
ADCET, Ashta, Sangli, India

diagnosis have simplified the task of a radiologist for manual detection of the tumors. Advancement in technology like artificial intelligence and machine learning has made diagnosis more accurate, time efficient, and cost-effective. In this paper, we discuss some of the machine learning techniques used for Brain tumor detection.

1.1 Challenges in Machine Learning

The kind of learning where a machine improves its performance based on experience is called as inductive learning. An agent is said to learn from experience with respect to some task and is measured by performance say 'p,' then 'p' improves with experience, and this is the actual concept of machine learning (a canonical definition stated by Tom Mitchell in 1997) [1]. There are several challenges to build a machine learning solution. These are: How good the model is? How one does choose a model? Does one have enough data? Is the data of sufficient quality? Is there any error in the data? Is one describing the data correctly? How confident can one be the result? The above questions have to be solved correctly in order to build a machine learning solution using some of the important machine learning assumptions/rules.

Supervised Learning: In the supervised learning, there are labeled data and the goal here is to come up with a function where input is given to get the output. The process for the supervised learning algorithm is shown in Fig. 1.

In the supervised learning, the training algorithm gets the input and produces the output and compares the output with the target, and if there is difference between the target and the output, error occurs which is sent to the agent for updating the

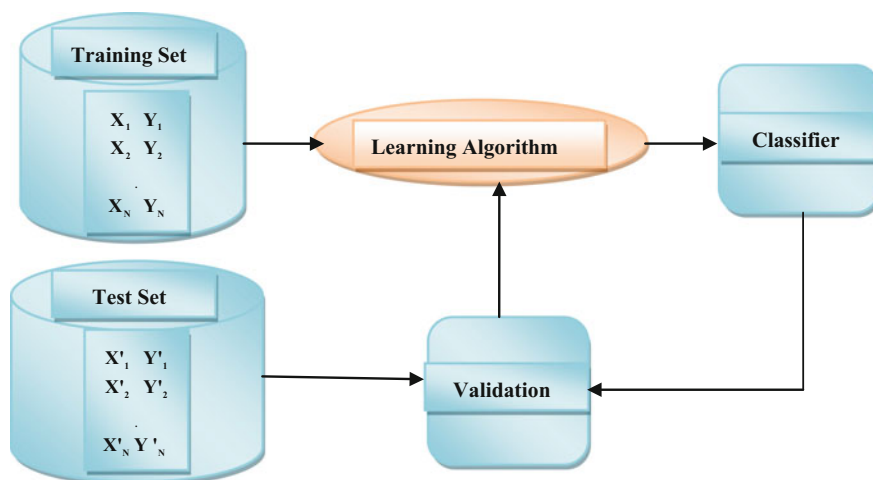


Fig. 1 Block diagram for process of supervised learning

weights [2]. Another problem in supervised learning is prediction or regression where the output predicted is not a discrete value and is a continuous value where most of the error occurs and the solution to fit the noise/error in the data is to over fit the data; here, care must be taken to avoid the over fit of the training data and so we can go for linear regression to minimize the sum of square errors where the aim is to minimize the prediction error by taking the square of the errors.

Unsupervised Learning: Here, there is an unlabeled training data set where clustering is done with groups of coherent or cohesive data points in the input space. The analysis is based on how spread out the points is that belong to the single group. In image processing, it can be with respect to image pixels depending on the similar pixels. The association rules are mining or learning frequent patterns or rules which are conditional dependency and these patterns are with respect to sequences, transactions, and graphs [2] (Fig. 2).

Reinforcement Learning: It is the learning of a machine to control a system through trial and error and minimal feedback (Fig. 3).

1.2 A Broad View on Different Techniques Used for Brain Tumor Detection

Brain tumor is detected from images acquired from Brain Magnetic Resonance Imaging (MRI). The image thus captured is preprocessed and then the region of interest is extracted; this process is called as segmentation of the image. The region of interest (ROI), i.e., the tumor region, is obtained by segmenting the input image which is a machine learning process.

The approach of the segmentation depends upon classifying the similar category of intensity values in an image and to form group of such similar intensity values (subjects). Here, the approach is based on two standards; i.e., similarity of intensity

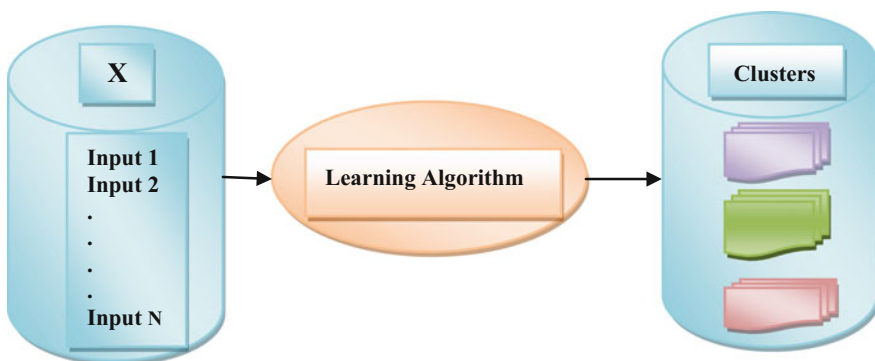


Fig. 2 Block diagram for process of unsupervised learning [1]



Fig. 3 Block diagram for process of reinforcement learning [1]

values is the dissimilarity between them and the task is either to find the similarity measure between the subjects and group them, and these subjects will have most similar index grouped into clusters or the dissimilarity among the objects separate to give the most dissimilar subjects in the space [3]. Figure 4 shows the different methods used for segmentation [4].

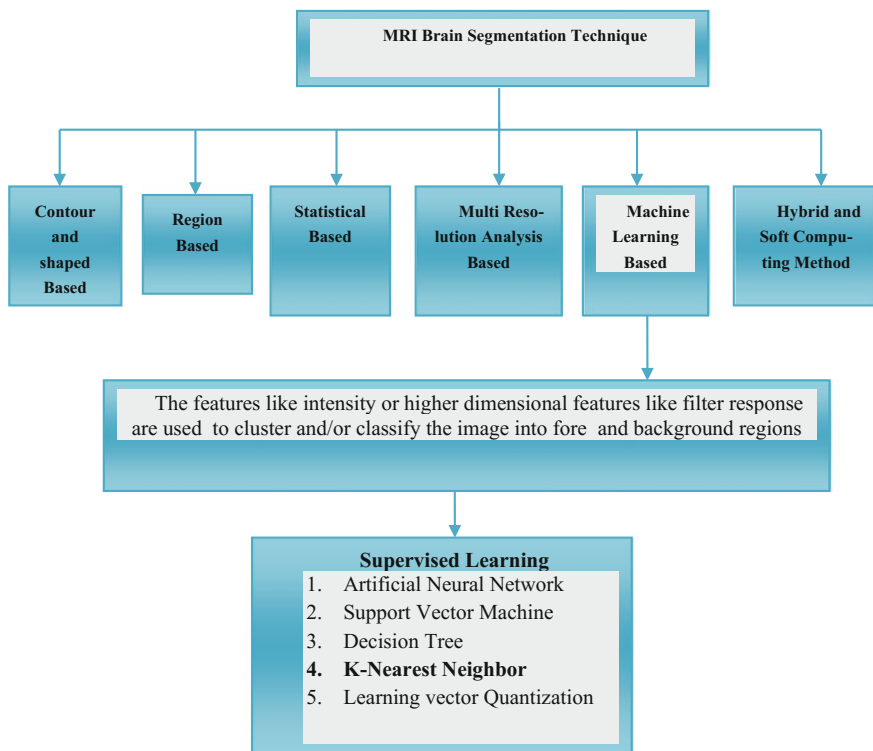


Fig. 4 Block diagram to represent a broad view on the techniques present for detection of Brain tumor

2 Related Work

In this paper, the segmented tumor is obtained by machine learning algorithm. The flow for the algorithm is as shown in the shaded region of Fig. 4. The Brain MRI images obtained are corrupt with noise so the initial step is to preprocess the image to reduce the unwanted noise and the high-frequency components. For a machine learning algorithm, the data is given as some kind of a probability distribution, where the input is drawn from a P-dimensional real space and the output is also a real space which can be given as

$$\text{Input } X = R^P \tag{1}$$

$$\text{Output } Y = R \tag{2}$$

The probability distribution is obtained from a set of samples that are drawn from $\{(x_1, y_1) (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)\}$, and these can also be called as training data set.

Learn a function

$$f(x)R^P = R \tag{3}$$

Then, the function produces a predicted output which can be given as

$$\hat{Y} = f(x) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p \tag{4}$$

$$f(x) = \beta_0 + \sum_{j=1}^p (\beta_j, x_j) \tag{5}$$

Set $X_0 = 1$ then

$$f(x) = \sum_{j=1}^p (x_j, \beta_j) \tag{6}$$

Equation (6) is of linear regression.

For the nearest neighbor classifier, we get

$$\hat{Y}(x) = \frac{1}{k_0} \sum_{xi=Nx}^{\infty} (Y) \tag{7}$$

The performance measure is to compare the true output with the predicted output where the loss function is given by

$$L = (Y, f(x)) \tag{8}$$

Find $f(x)$ to minimize the loss function where the loss function is the expected predicted error. Minimal error is obtained on conditioning a specific value of x

$$f(x) = \arg \min E_y |x| (y - c)^2 |x \tag{9}$$

The value of C is the error and should be as small as possible, and this is called as the conditioning on a point.

$$\therefore (y - x)^2 = EPE(x) \tag{10}$$

One of the multi-clustering algorithms that are used for extracting tumor in MR images is the k-means clustering. This is a region-based segmentation which divides the region into number of clusters. The required number of clusters is represented by ‘K’ for particular application. Here, the technique is to find the nearest distance of every pixel to each cluster. Compared with the traditional algorithms like hierarchical algorithms, one of the advantages of using k-means algorithm is that for huge sample space on smaller value of ‘k’ the algorithm runs much faster. The other limitations that are associated with the algorithm are: 1. For a certain application, it is pretty difficult to guess the ideal value of ‘k’; 2. Bounding of the initial partitioning varies these results in different clusters; 3. One of the main factors is the density of the clusters which affects the performance of the algorithm as the clusters with different density algorithm does not work well enough [5].

Similarly K-Nearest Neighbor (K-NN) is a clustering algorithm, which is proficient in integration the nearby pixels that have similar intensity values by measuring the Euclidean distance among the classes of K-samples of the Brain MRI. The algorithm is shown in Fig. 5.

The MRI images are first preprocessed using some of the image enhancement techniques like filtering, mostly the apostrophic filtering or the median filtering techniques are used. The main component as an input to the K-NN algorithm is the

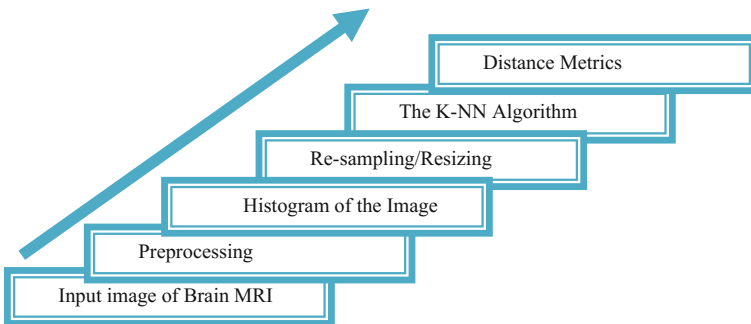


Fig. 5 Block diagram to represent of the K-NN algorithm the flow shows the bottom-up approach

feature values of various intensity levels obtained by the histogram of the input. The next step is re-sampling where the proper geometrical representation of the image is obtained [6]. The K-NN is a non-parametric algorithm which is trained with data, as the number of trained samples increases the output is more efficient or accurate. If S is the number of samples and C is the number of classes where $S > C$, the distance between the two pixels is found out by the Euclidean distance measure which is given as [7]. Let $X = (x_1, x_2, x_3, \dots, x_n)$ and $U = (u_1, u_2, u_3, \dots, u_n)$ be the two points then Y , the distance of order is defined as

$$D(x, u) = \sum_{i=1}^n (|x_i - u_i|^y)^{\frac{1}{y}} \quad (11)$$

- For $y = 1$: Manhattan Distance
- $y = 2$: Euclidean Distance
- $y = \infty$: Infinity Distance

Thus, the distance metrics are obtained and the results for some of the input images are as given in Table 2.

3 Experimental Results

Table 1 shows the results of the segmented tumor, and the corrupted input MRI images are preprocessed to initially enhance and eliminate the unwanted noise. High-frequency components are extracted from the enhanced image and the segmented output is obtained by clustering using the k-means algorithm. Manhattan distance metric is used for distance measurement. The segmented output is as shown in the last column of Table 1.

One of the supervised learning algorithms is the k-nearest neighbor. Researcher (Sudharani) [7] worked on a set of images to train as well as test the K-NN algorithm. The results obtained are shown in Tables 2 and 3 where Table 2 gives the distance measure of the data of tumors and Table 3 gives classification and identification score.

Table 1 Results for different input images using k-means algorithm

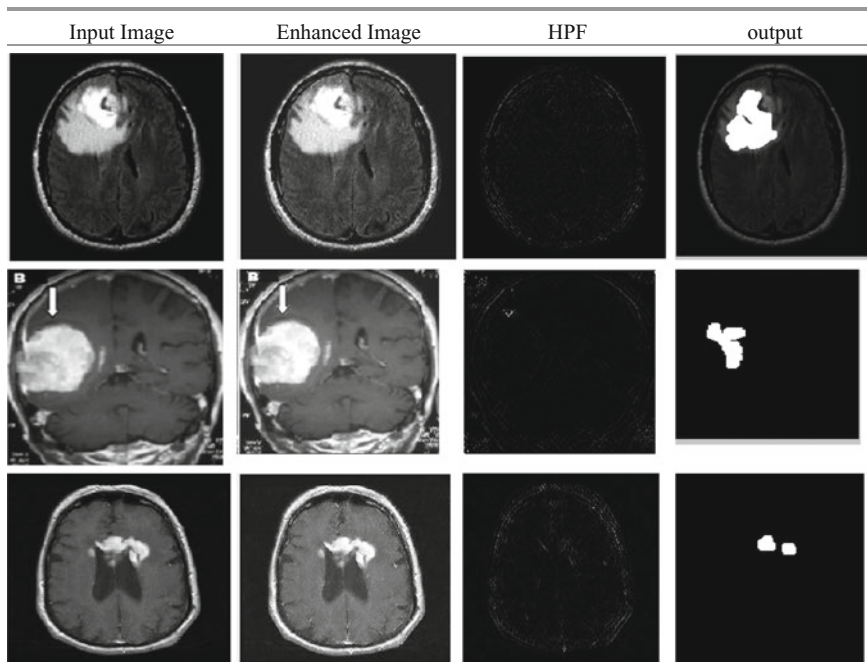


Table 2 Results for distance metrics

S. No.	Tumor 1	Tumor 2	Tumor 3	Tumor 4
Tumor1	0	53.57	312.43	47.67
Tumor2	53.57	0	274.71	73.22
Tumor3	312.43	274.71	0	328.11
Tumor4	47.67	73.22	328.11	0

Table 3 Results for different input images using K-NN algorithm. The classification and the identification score

S. No.	Actual label	Classification score	Identification score
Patient 1	Tumor 1	252.76	992.29
Patient 2	Tumor 2	178.78	883.99
Patient 3	Tumor 3	506.34	977.06
Patient 4	Tumor 4	446.7	969.29

4 Conclusion

This research is piloted to identify Brain tumor using medical imaging techniques. Here, more emphasis is given on the machine learning techniques and the results obtained are better than the results got from morphological operations. We observe that machine learning techniques have the learning capability which adapts itself for new data sets, which helps in minimizing errors by updating weight. Taking into account these powerful techniques of machine learning we are looking forward to find the volume of the tumor which is even more challenging which may help the radiologist for planning the therapy of diagnosis.

References

1. Chapelle, O., Schölkopf, B., Zien, A. (eds.): *Semi-Supervised Learning*, pp. 508. MIT Press, London, U.K (2006). ISBN:978-0-262-03358-9
2. Duch, W., Mańdziuk, J. (eds.): *Challenges for computational intelligence*. In: *Series on Studies in Computational Intelligence*, Vol. 63, pp. 488. Springer, New York (2007). ISBN:978-3-540-71983-0
3. Soleimani, V., Vincheh, F.: Improving ant colony optimization for brain MR image segmentation and brain tumor diagnosis. In: *First Iranian Conference on Pattern Recognition and Image Analysis (PRIA)*. IEEE (2013)
4. El-Dahshan, E.-S.A., Mohsen, H.M., Revett, K., Salem, A.-B.M.: Computer-aided diagnosis of human brain tumor through mri: a survey and a new algorithm. *Expert Syst. Appl.* **4**, 5526–5545 (2014), Contents lists available at Science Direct. www.elsevier.com/locate/eswa. <https://doi.org/10.1016/j.eswa.2014.01.021>
5. Khare, S., Gupta, N., Srivastava, V.: Optimization technique, curve fitting and machine learning used to detect brain tumor in MRI. In: *2014 IEEE International Conference on Computer Communication and Systems (ICCCS 114)*, 20–21 Feb 2014, Chennai, India. <https://doi.org/10.1109/icccs.2014.7068202>
6. Selvakumar, J.: Brain tumor segmentation and its area calculation in brain MR images using K-mean clustering and Fuzzy C-mean algorithm. In: *IEEE Conference on ICAMSE*, pp. 186–190 Mar 2012
7. Sudharanil, K., Sarma, T.C., Satya Rasad, K.: Intelligent brain tumor lesion classification and identification from MRI images using kNN technique. In: *2015 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*. 978-1-4673-9825-1/15/\$3 1.00 ©2015 IEEE. <https://doi.org/10.1109/iccicct.2015.7475384>