

# Framework on an Intelligent Ensemble Technique for Brain Tumor Detection and Classification

Abiodun Ipaye<sup>1</sup>, Prof. O. O. Obe<sup>2</sup>, Prof. Thompson<sup>3</sup>

<sup>1</sup>Department of Computer Science, Federal College of Agriculture Akure, Ondo State, Nigeria

<sup>2,3</sup>Department of Computer Science, Federal University of Technology Akure, Ondo State, Nigeria

DOI: <https://doi.org/10.5281/zenodo.7769192>

Published Date: 25-March-2023

---

**Abstract:** A tumor is one of the most dangerous malignancies that may affect human beings. The traditional approach of recognition and categorization of brain tumors by radiologists or clinical experts is by human inspection of medical resonant images (MRI). This type of approach is time-consuming and tedious and most of the time the result is not accurate. When handling human life, accuracy is highly required, hence the need for the automation of brain tumor detection and classification with better accuracy. In the past, various researchers have proposed different traditional classifiers for detecting brain tumors. In this research paper, our model intends to increase the effectiveness of traditional machine learning algorithms by using the stacking method to combine a set of support vector machines, decision trees, and k-nearest neighbors. These base learners are proposed because of their low computational time complexity. The proposed methodology is divided into modules. Module 1: MRI Image dataset Acquisition from Kaggle.com, Module 2: Preprocessing and Extraction of Features using GLCM, Module 3: ensemble approach for classification, and finally, Module4 is the evaluation of the results from the classifiers using standard performance evaluation metrics

**Keywords:** Ensemble, benign, malignant, support vector machines, decision trees, and k-nearest neighbors.

---

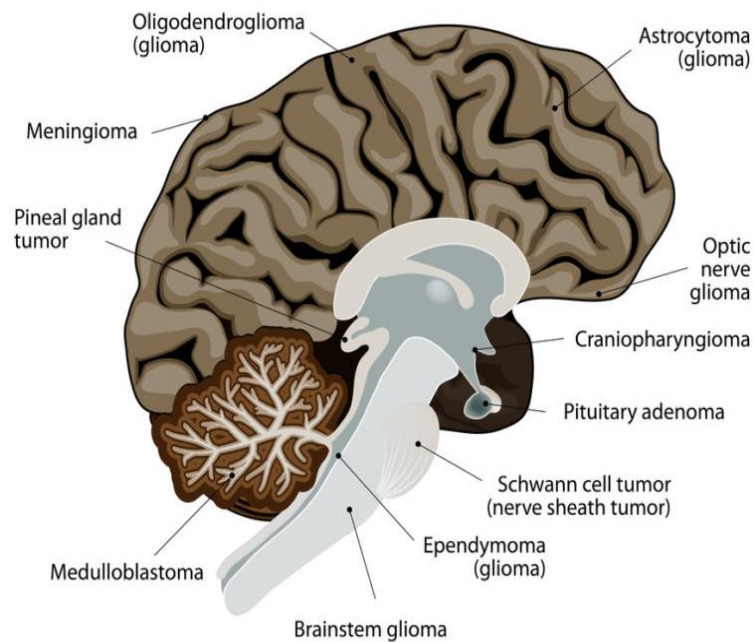
## 1. INTRODUCTION

The human brain is a powerful and interesting organ that is essential to our day-to-day existence. It serves as the focal point of our conscious experience and regulates all of our ideas, deeds, intelligence, speech, memory, senses, thoughts, physical activity, taste, creativity, and feelings. More than 100 billion neurons make up the brain, which is connected by trillions of synapses and functions as a sophisticated network that processes and stores information. Therefore, any accident or damage to this important organ will impair the body's normal operation and lead to irregular routines. Therefore, it is essential to treat this organ with the highest respect. The most severe and possibly deadly disorder affecting the brain nowadays is a brain tumor.

The two basic types of brain cancers are primary and secondary brain tumors. While secondary brain tumors originate as cancer elsewhere in the body and spread to the brain, primary brain tumors develop within the brain. Brain Tumor Symptoms include Morning headache or headache that goes after throwing up, Seizures, Vision, hearing, and speech issues, decrease in appetite, frequent vomiting and nausea personality, mood, concentration, or behavior changes, difficulty walking and losing one's equilibrium Weakness, abnormal slumber or change in activity level. Etc. Environmental risk factors, such as exposure to toxic chemicals, may be one of the causes of brain tumors. Additionally, they may be hereditary, such as having a defect from birth or developing helplessness from one's parents. As one age, these inherited mutations and susceptibilities may also compound with time.

When handling human life, accuracy is highly required, hence the need for the automation of brain tumor detection and classification. This paper proposes the implementation of an intelligent ensemble technique for brain tumor detection and classification of Magnetic Resonance Imaging (MRI) images using the stacking ensemble technique. The model proposed the combination of support vector machine, decision tree, and k-nearest neighbor as the base learners while logistic regression will be used as the meta-model for the final prediction. The specific objectives of the research are to:

- i. design a brain tumor detection and classification model using ensemble technique,
- ii. implement the model in (i), and
- iii. assess the effectiveness of the model using standard evaluation metrics.



**Figure 1: The most common primary brain tumors. ([https://drstegall.com/wp-content/uploads/2017/05/brain\\_tumor-862x793.jpg](https://drstegall.com/wp-content/uploads/2017/05/brain_tumor-862x793.jpg))**

## 2. RELATED WORK

In the work of Basir et al. (2019), They tried to categorize brain tumors as benign or malignant, and they estimated the tumor region's size. For segmentation, Otsu's Threshold technique was employed. Thirteen characteristics were extracted for categorization using the Stationary Wavelet Transform (SWT), Principal Component Analysis (PCA), and Gray Level Co-occurrence Matrix (GLCM). Based on the Majority Voting approach, a hybrid ensemble classifier (KNN-RF-DT) performs the classification. The project tries to improve the performance of traditional classifiers rather than deep learning. The approach was evaluated on a dataset of 2,556 pictures, which were split up 85:15 for training and testing, and it produced accurate results of 97.305%.

In the 2015 paper titled "Brain Tumor Identification and Segmentation" by *Megersa et al.*, a method for recognizing, segmenting, and stripping the skull from brain images was proposed. The method involves using a Hybrid Intelligent Algorithm that combines the Fuzzy Hopfield neural network and the traditional FCM algorithm. The Hopfield Neural Network approach uses the Lyapunov energy function to optimize problems, which differs from the standard FCM technique that minimizes an objective function based on least squares. To automatically choose the threshold value, Otsu's method was used while local entropy reduction with a bicubic spline model was applied to handle intensity inhomogeneity in the brain MR image. The proposed model was validated using T1-w and T2-w head MRI scans as the source of data. Three distinct datasets were used to confirm the efficacy of the suggested system, including benign MR images from the Internet Brain Segmentation Repository (IBSR), simulated tumor MR images, and malignant brain MRI scans from Pioneer Diagnostic Center (PDC), Addis Abeba, Ethiopia. The primary objective of the suggested framework is to build a model capable of separating normal tissues and can be applied in real-time scenarios. The results obtained from the different datasets are as follows: For the PDC dataset, the sensitivity and accuracy were 0.9273 and 81%, respectively. The simulated

dataset produced sensitivity and accuracy of 0.8244 and 82%, respectively. When the S08 dataset, where the detection of tumorous slices based on symmetry fails, is removed, the sensitivity increases to 92%, while the specificity decreases to 80%.

Ramya P. *et al.* (2021) In the study Brain tumor segmentation utilizing cluster ensemble and deep super learner for MRI classification, self-organization map (SOM) and the ensemble of K-means, gaussian mixture model, and self-organization map, and their performance are reviewed. As a consequence, Ensemble1 (K-means, SOM, GMM) produced results with 96.3% accuracy whereas Ensemble2 (K-means, SOM, FCM) produced results with 95.8% accuracy. In comparison to the K-means, SOM, and FCM technique, the combination of K-Means, SOM, and GMM produced greater segmentation accuracy. The main objective of the research is to introduce a novel ensemble model that makes use of two sets of four different clustering techniques.

Kang *et al.* (2021) proposed an MRI-Based Brain Tumor Classification Using an Ensemble of Deep Features and Machine Learning Classifiers (ResNet-50, ResNet-101, DenseNet-121, VGG-16, VGG-19, AlexNet, Inception V1 (GoogLeNet), Inception V3, and MobileNet V2). They did this by using a variety of pre-trained deep convolutional neural networks as feature extractors from the MRI images. To distinguish between normal and pathological brain MR images, various ML classifiers were used, including neural network classifiers, Naive Bayes classifiers, AdaBoost classifiers, k-NN classifiers, RF classifiers, SVM classifiers, and ELM classifiers. The BT-small-2c (benign/malignant), BT-large-2c (benign/malignant), and BT-large-4c make up the dataset. The results of the work that has been presented show that SVM frequently performs better than other ML techniques. The research's primary objective is to suggest a novel model for tumor identification. In conclusion, further work needs to be done to lower the size of the model.

The research of Brunese *et al.* (2019) titled an Ensemble Learning Approach for Brain Cancer Detection exploiting Radiomic Features. Various machine learning algorithm was used to perform Classification. The best result from various classifiers is obtained through voting. The work aims to classify MRI images into grade I, II, III, or IV brain tumor. The research obtained the following results: precision of 99% (I grade), 99% (II grade), 98% (III grade), and 99% (IV grade), 99.5% recall (I grade), 98% (II grade), 99.5% (III grade) and 98.9% (IV grade).

According to Chander *et al.* (2020), brain tumors pose a hazard to people. Therefore, early tumor diagnosis is crucial before administering the appropriate treatment. Using an MRI image as input, for segmentation, the adaptive K-Means clustering technique was proposed, and support vector machines were used for classification. By using this approach, a 99.7% segmentation accuracy is attained. Support vector machines can classify objects with an accuracy of 93%. To increase contrast and fill in any gaps in the image, the intensity of the image is changed. Building a classification model with three kernels—linear, quadratic, and polynomial—and comparing their accuracy is the aim of this study.

Sabitha *et al.* (2021) introduced principal component analysis (KPCA) and support vector machine (KSVM). While KSVM was utilized for the classifier, KPCA is used to reduce the extracted features to essential features. The proposed work was segmented into four stages which include: Pre-processing and Segmentation, Feature extraction with feature reduction, and Classification. Over 90% of the result is accurate.

Singh *et al.* proposed the use of K-Means Segmentation and the Normalized Histogram method for the effective identification of tumors from MRI in their 2016 study. The objective of this technique is to represent the image in a more meaningful and useful manner that is also easier to analyze. K-means segmentation is an unsupervised method that separates a set of data points into k clusters. The Median filter, Adaptive filter, Averaging filter, Un-sharp masking filter, and Gaussian filter are used to eliminate noise during the preprocessing stage. The goal of this suggested model is to create a novel model that can predict the precise location of a brain tumor. The accuracy result achieved from the proposed model is SVM 91.49% and Naïve Bayes 87.23%. However, the model has some restrictions, such as poor outcomes. It was impossible to determine the actual site of a few of the tumors, and the picture quality was poor. As a result, a better algorithm may be employed in the future for improved outcomes.

Because there are so many MRI images available to medical professionals, classifying and detecting brain tumors is a highly laborious process, and occasionally noise in the MRI scans causes the results to be unreliable, hence the need for automation. Shree *et al.* (2018) recommended feature extraction, DWT, and probabilistic neural networks for the identification and classification of MRI images of brain tumors. It was possible to achieve 100% accuracy for test data and 95% accuracy for training data. The accuracy of the outcome could be improved in the future. The isolation, detection, and extraction of the infected area on magnetic resonance imaging (MR) is a major but tedious but time-consuming task performed by

radiologists or clinical specialists, and their accuracy depends solely on their experience. Therefore, the use of computer-assisted technology becomes even more necessary to overcome these limitations. Image Analysis for MRI-Based Brain Tumor Identification and Feature Extraction is a 2017 study by Bahadure *et al.* It was possible to reach accuracy results of 96.51%, 94.2% specificity, and 97.72% sensitivity utilizing biologically inspired BWT and SVM. The model's performance was compared with that of ANFIS, Back Propagation, and K-NN classifier. The outcome could still be enhanced.

A cutting-edge deep neural network (DNN) was built by Ullah *et al.* (2020) to distinguish between benign and malignant brain tumors in MRI scans. The method, which had a 95.8% accuracy rate, included preprocessing, feature extraction using a 2D discrete Wavelet Transform, feature reduction using color moments, and classification using a feed-forward neural network. The study objective is to put forth a model that can automatically identify brain tumors from MRI scans, assisting doctors in making decisions quickly as the human procedure is time-consuming and ineffective. The work's shortcoming is that the model needs to be retrained anytime the MRI database grows. Therefore, a larger dataset including a variety of virus types should be used to train the model.

### 3. METHODOLOGY

The proposed methodology has been divided into modules as follows:

- i. Module 1: MRI Image dataset Acquisition from Kaggle.com
- ii. Module 2: Preprocessing and Feature Extraction
- iii. Module 3: Classification
- iv. Module 4: Evaluation

#### Module 1: MRI Image dataset Acquisition

The dataset for the model will be acquired online from Kaggle.com. Where the dataset obtained online is less than the proposed number for the study, the data Augmentation technique will be used to increase the number of MRI images to the 10,000 proposed for the research. The dataset contains MRI images that are benign and malignant.

**Data Augmentation:** The dataset will be enhanced by several random modifications using this approach to expand its size. The possibility of over-fitting will decrease with data augmentation, and the model will be better trained to generalize. horizontal and vertical flips are 0.1. Figure 2 below is an illustration of data augmentation

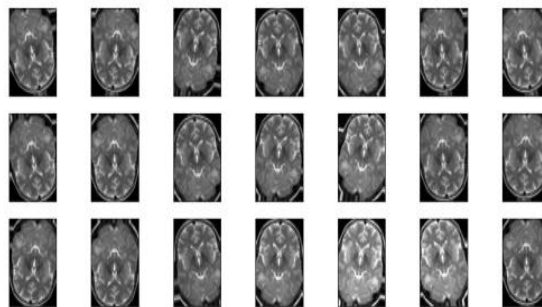


Figure 2: Augmented Images

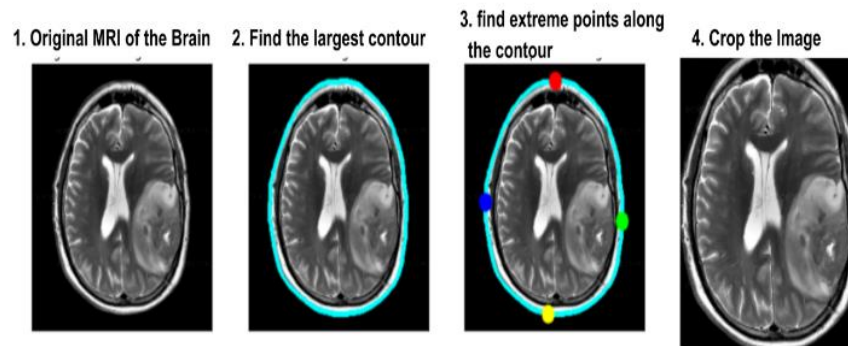
Source: Donghyun Kim (2021)

#### Module 2: Preprocessing and Feature Extraction

The three main tasks in Module 2 are: dividing the dataset into ratios; cropping or trimming the images and resizing the images. The main aim of this phase is for the model to learn accurately.

**Dataset Splitting:** 10,000 MRI images will be used in the dataset for the proposed study's model. The dataset will be split into a training dataset and a testing dataset, with 30% of the dataset used for testing and 70% of the dataset used for model training.

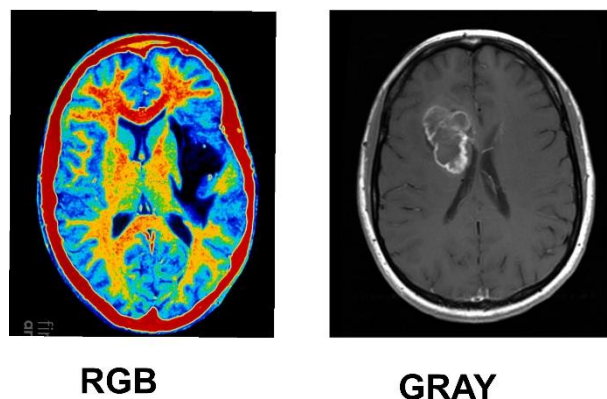
**Image Cropping:** The following procedures will be used to remove the brain portion from the MRI image: locates the broadest portion of the brain's contour, locates its most extreme points, and crops the image. Figure 3 below describes this procedure:



**Figure 3: Cropped Image**

Source: Donghyun Kim (2021)

**Image Resizing:** Because the dataset contains images with various dimensions and aspect ratios, the images will be shrunk to fit the input size of the pre-trained model. A consistent dimension of 224 224 3pixel will be utilized. The RGB image will be converted to a gray scale using the RGB2GRAY scale function.



**Figure 4: RGB to GRAY image**

### Module 3: Classification

An ensemble of a decision tree, k-nearest neighbors, and support vector machines are the proposed classification methods that will be applied in the model. By building multiple models and merging them, Ensemble finds a solution to a problem. Bagging, boosting, and stacking are three most common examples of ensemble approaches that are used for categorization. We'll employ stacking in this study.

**Stacking:** Following training on the entire dataset, the base models generate features (as output), which are then utilized to train the meta-model. In stacking, the fundamental models are frequently different. Finding characteristics from base models with the greatest accuracy is made easier by the meta-model.

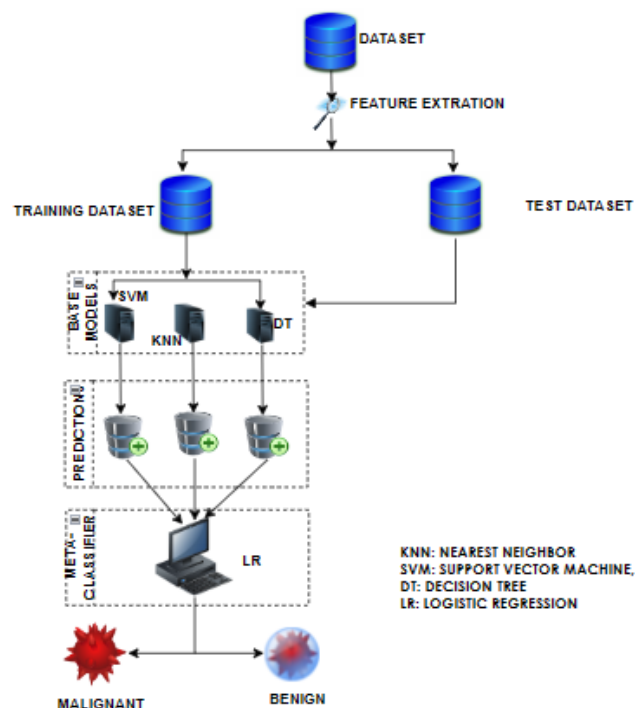
#### Algorithm:

1. Start by dividing the training dataset into n sections.
2. After fitting a fundamental model to n-1 parts, predictions are generated for the nth component. This is done for each of the n parts of the train set.
3. The fundamental model is then fitted to the complete train dataset.
4. The test dataset is predicted using this model.

5. Repeating steps 2 through 4 for a different base model yields a different set of predictions for the train and test dataset.
6. The training data set predictions are utilized as a feature to create the new model.
7. With the testing dataset and this final model, predictions are made.

#### Benefits of a Stacked Generalization Model include:

1. One advantage of stacking is that it may pool the abilities of many highly effective models to produce predictions that surpass any individual model in the ensemble on a classification or regression task.
2. Stacking enhances the model prediction accuracy.



**Figure 5: Stack ensemble Architecture**

In the above architecture, we can see that we have a training dataset that is trained on different classifiers SVM, decision tree, and KNN which give the result **P1**, **P2**, and **P3**, we can take all predictions with help of **Stacking** and we got the final result **Pf**.

#### Classifiers

According to Hosni *et al.* (2019) in their work Computer Methods and Programs in Biomedicine they investigated 40 publications (i.e., studies proposing heterogeneous ensembles, 64 ensembles were used. The mapping investigation discovered 16 different classifiers that were utilized to create the heterogeneous ensembles. Five of the classifiers were used over 20 times.

In fact, in 42 ensemble combinations, SVM was the approach that was most frequently utilized, followed by DT in 38, the K-NN method in 33, ANN in 31, and BC 25 combinations, respectively. Hence, the choice of SVM, DT, and K-NN as the base learning.

#### Support Vector Machine

A support vector machine represents a data set as points in space divided into categories by a distinct gap or line that stretches as far as feasible. The additional data points are now mapped into the same area and categorized into one of many categories based on which side of the line or separation they land on.

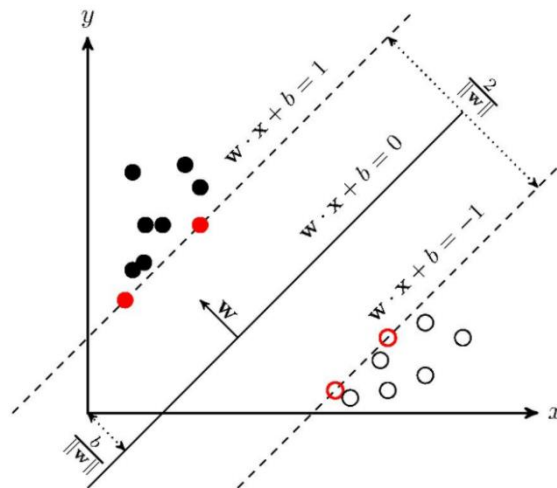


Figure 6: Graph showing hyperplane (Source: Ray 2017)

By calculating the decision function  $w^T x + b = w_1 x_1 + \dots + w_n x_n + b$ , the linear SVM classifier model predicts the class of a new instance  $x$ : if the result is positive, the predicted class  $y_i$  is Benign (1), else it is the Malignant class (0);

$$y_i = \begin{cases} 1 & \text{if } w^T x + b \geq 1 \\ 0 & \text{if } w^T x + b \leq -1 \end{cases}$$

where:

$w$  = weight vector

$b$  = bias

$x$  symbolizes the training examples closest to the hyperplane.

$T$  is transpose

### Decision Tree

Decision trees can be used for both classification and regression since they can handle both numerical and categorical data. The data set is divided up into increasingly smaller pieces, or nodes, as the tree gets bigger. The output of a decision tree consists of decision and leaf nodes, each of which represents a decision and comprises two or more branches. The topmost node that corresponds to the best predictor is known as the root node.

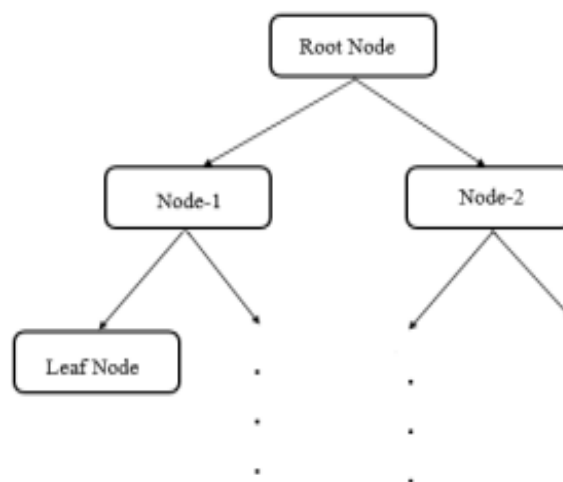


Figure 7: Basic Structure of Decision Tree (Source: Chergui 2019)

To build our classification tree we use Entropy (H)

$$\blacktriangleright H(S) = -\sum_i^C P_x \log P_x$$

- Set of all instances, S
- N= number of unique class values.
- P<sub>x</sub> - Probability of the event
- C - Number of classes

$$\text{Gain}(S, F) = \text{Entropy}(S) - \sum(|S_v|/|S|) * \text{Entropy}(S_v)$$

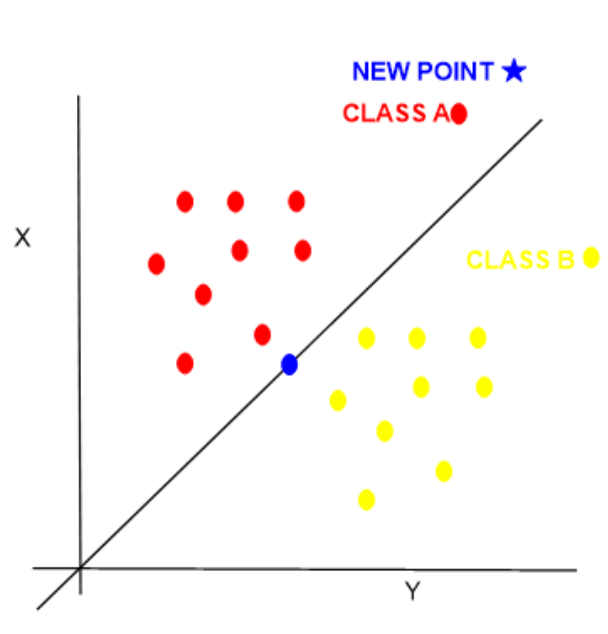
where:

S is the set of samples being split, F is the feature used for splitting, S<sub>v</sub> is the subset of samples that have the value v for feature F, |S<sub>v</sub>| is the number of samples in subset S<sub>v</sub> |S| is the total number of samples in set S, Entropy(S) is the entropy of set S, Entropy (S<sub>v</sub>) is the entropy of subset S<sub>v</sub>

### K-nearest neighbors (KNN)

The K-nearest neighbors (KNN) method uses the concept of "feature similarity" or "nearest neighbors" to forecast which cluster a new data point will belong to. The stages listed below will help us understand how this algorithm functions better.

Step 1 involves preparing a dataset that is free of errors and suitable for use with machine learning techniques. Assuming that we have already cleaned the dataset and separated it into training and testing sets, we can move on to Step 2. In this step, we need to decide on the value of K, an integer that determines the number of nearby data points used by the algorithm. Step 3 is a repetitive process that applies to every data point in the dataset.



**Figure 8: Example of k-NN classification. The blue dot should be classified either to yellow dot or to red dot**

The two most frequently utilized methods for determining the proximity between a new data instance and the existing training data instances are the Manhattan distance and the Euclidean distance. For our research, we chose to employ the Euclidean distance measure for the k-NN algorithm. To compute the Euclidean distance  $d$  between two data points  $x$  and  $y$ , the following formula will be used:

$$d(x, y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$$



The formula refers to finding the distance (d) between two points (x,y) with N dimensions (attributes). The i-th attribute (component or data object) of x and y are represented by  $x_i$  and  $y_i$ .

### Meta-Classifer

The Stacking ensemble learning approach utilizes a meta-classifier to merge various classification models. In our model, the suggested meta-classifier is logistic regression, which is a well-known multivariate technique used in biomedical informatics. Logistic regression is frequently used for prediction and has been successfully applied in cancer prediction studies (Samatha, 2009).

The formular for logistic regression can be written as:

$$\text{Log}\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

### Module 4: Evaluation

The model will be evaluated using standard performance metrics

### EXPECTED CONTRIBUTION TO KNOWLEDGE

At the end of this research, an intelligent ensembled model for brain tumor detection and classification would have been developed to aid relevant medical practitioner in brain tumor diagnosis for treatment

## 4. CONCLUSION

In the field of medicine, it can be time-consuming for an expert to manually identify a brain tumor from an MRI image, and this approach may not be suitable for handling a large amount of data. Instead of relying on manual identification, image processing and machine learning techniques can be employed to detect tumors from the images.

### REFERENCES

- [1] National Cancer Institute, Cancer.gov Updated: August 6, 2021
- [2] American Brain Tumor Association (ABTA 2015) about brain tumors, a primer for patients and caregivers, [www.abta.org](http://www.abta.org)
- [3] G. Garg and R. Garg. (2021) Brain Tumor Detection and Classification based on Hybrid Ensemble Classifier. ArXiv, 2021, volume abs/2101.00216
- [4] Y. Megersa and G. Alemu, "Brain tumor detection and segmentation using hybrid intelligent algorithms," AFRICON 2015, 2015, pp. 1-8, doi: 10.1109/AFRCON.2015.7331938.
- [5] P. Ramya, M.S. Thanabal, & C. Dharmaraja, (2021). Brain tumor segmentation using cluster ensemble and deep super learner for classification of MRI. Journal of Ambient Intelligence and Humanized Computing, 12(10), 9939–9952. doi:10.1007/s12652-021-03390-8
- [6] J. Kang, Z. Ullah, & J. Gwak, (2021). MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers. Sensors, 21(6), 2222. doi:10.3390/s21062222
- [7] L. Brunese, F. Mercaldo, A. Reginelli, & A. Santone, (2019). An Ensemble Learning Approach for Brain Cancer Detection exploiting Radiomic Features. Computer Methods and Programs in Biomedicine, 105134. PII: S0169-2607(19)30511-5, doi:10.1016/j.cmpb.2019.105134
- [8] C.P. Sharath, J. Soundarya, & R. Priyadharsini, (2019). Brain Tumour Detection and Classification Using K-Means Clustering and SVM Classifier. RITA 2018, 49–63. doi:10.1007/978-981-13-8323-6\_5
- [9] V. Sabitha, J. Nayak, & P.R. Reddy, (2021). MRI brain tumor detection and classification using KPCA and KSVM. Materials Today: Proceedings Available online 4 March 2021. doi:10.1016/j.matpr.2021.01.714
- [10] G. Singh, & M.A. Ansari, (2016). Efficient detection of brain tumor from MRIs using K-means segmentation and normalized histogram. 2016 1st India International Conference on Information Processing (IICIP). doi:10.1109/iicip.2016.7975365

- [11] N.B. Bahadure, A.K. Ray, & H. P. Thethi, (2017). Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM. *International Journal of Biomedical Imaging*, 2017, 1–12. doi:10.1155/2017/9749108
- [12] Z. Ullah, M.U. Farooq, S.H. Lee, & D. An, (2020). A Hybrid Image Enhancement Based Brain MRI Images Classification Technique. *Medical Hypotheses*, 109922. doi:10.1016/j.mehy.2020.109922
- [13] Z. U. Rehman, S.S. Naqvi, T.M. Khan, M.A. Khan, T. Bashir. (2019). Fully automated multi-parametric brain tumour segmentation using superpixel based classification. *Expert Syst Appl* 2019;118:598–613.
- [14] F. O. Akinloye, O. O. Obe, and O. K. Boyinbode (2020). Development of an affective-based e-healthcare system for autistic children. *Elsevier Science Direct, Scientific African*, Volume 9 (1): 1-9.
- [15] Y. A. Asogbon, W. S. Oluwarotimi., E. Nsugbe, Y. A. Jarrah, O. O. Obe, Y. Geng, and G. Li (2021). A Deep Learning-based Model for Decoding Motion Intent of Traumatic Brain Injured Patients' using HD sEMG Recordings. 2021 IEEE International Workshop on Metrology for Industry 4.0 and IoT.
- [16] O.E. Oduntan, I.A. Adeyanju, A.S. Falohun, O.O. Obe, "A comparative analysis of euclidean distance and cosine similarity measure for automated essay-type grading *Journal of Engineering and Applied Sciences*, 2018, 13(11), pp. 4198–4204
- [17] C.C Ugwu, O.O. Obe, O.S. Popoola, A.O. Adetunmbi, "Adistriuted denial of service attack detection system using long short term memory with Singular Value Decomposition", *Proceedings of the 2020 IEEE 2nd International Conference on Cyberspace , CYBER NIGERIA 2020*, 2021, pp. 112-118, 9428870.
- [18] F. T. Matthew, A.I. Adepoju, O. Ayodele, O.O. Obe, O. Olaniyan, A. Esan, B. Omodunbi, F. Egbetola, "Development of Mobile-Interfaced Machine Learning-Based Predictive Models for Improving Students' Performance in Programming Courses" *International Journal of Advanced Computer Science and Applications (ijacsa)*, 9(5), 2018. <http://dx.doi.org/10.14569/IJACSA.2018.090514>