BCM-VEMT: Classification of Brain Cancer from MRI Images using Deep Learning and Ensemble of Machine Learning Techniques

Prottoy Saha (prottoy@cse.kuet.ac.bd)
Khulna University of Engineering and Technology
Rudra Das
Khulna University of Engineering and Technology
Shanta Kumar Das
Khulna University of Engineering and Technology

Research Article

Keywords: Brain Cancer, Convolutional Neural Network, Transfer Learning, Ensemble of Classifiers, Machine Learning, MRI Images

Posted Date: December 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1100868/v1

BCM-VEMT: Classification of Brain Cancer from MRI Images using Deep Learning and Ensemble of Machine Learning Techniques

Prottoy Saha, Rudra Das, Shanta Kumar Das

Department of Computer Science and Engineering

Khulna University of Engineering & Technology,

Khulna-9203, Bangladesh

prottoy@cse.kuet.ac.bd, das1607070@stud.kuet.ac.bd, das1607068@stud.kuet.ac.bd

Abstract: Brain Cancer is quite possibly the most driving reason for death in recent years. Appropriate diagnosis of the cancer type empowers the specialists to make the right choice of treatment, decision and to save the patient's life. It goes no saying the importance of a computer-aided diagnosis system with image processing that can classify the tumor types correctly. In this paper, an enhanced approach has been proposed, that can classify brain tumor types from Magnetic Resonance Images (MRI) using deep learning and an ensemble of Machine Learning Algorithms. The system named BCM-VEMT can classify among four different classes that consist of three categories of Brain Cancers (Glioma, Meningioma, and Pituitary) and Non-Cancerous which means Normal type. A Convolutional Neural Network is developed to extract deep features from the MRI images. Then these extracted deep features are fed into multi-class ML classifiers to classify among these cancer types. Finally, a weighted average ensemble of classifiers is used to achieve better performance by combining the results of each ML classifier. The dataset of the system has a total of 3787 MRI images of four classes. BCM-VEMT has achieved better performance with 97.90% accuracy for the Glioma class, 98.94% accuracy for the Meningioma class, 98.00% accuracy for the Normal class, 98.92% accuracy for the Pituitary class, and overall accuracy of 98.42%. BCM-VEMT can have a great significance in classifying Brain Cancer types.

Keywords: Brain Cancer, Convolutional Neural Network, Transfer Learning, Ensemble of Classifiers, Machine Learning, MRI Images.

Statements and Declarations

Competing Interests: The authors declare no competing interests.

1 Introduction

The tumor means swelling of a part of the body, normally without inflammation, due to abnormal growth of tissue, either benign or malignant [1]. A brain tumor implies the accumulation of unusual cells in certain tissues of the brain**.** The benign ones are not considered as cancerous and cannot be spread to different parts of the brain or body. The malignant ones are considered cancerous, grow uncontrollably, and can be spread to other parts of the body [2]. Every year, all over the world, 12.7 million people are diagnosed with cancer, and 7.6 million people died of cancer [3]. National Brain Tumor Society says that in the United States, 700,000 individuals are living with a primary brain tumor, and in 2021 around 85,000 more will be analyzed [4].

According to the National Brain Tumor Society, there are more than 120 types of brain tumors [5]. Among these types, Glioma, Meningioma, and Pituitary tumors are the most primary types. These three types cover almost 75% of all brain tumor types including 45% for Glioma type, 15% for Meningioma type, and 15% for Pituitary type [6].

Magnetic Resonance Imaging (MRI) is the most refined and waves used magnetic resonance imaging to obtain high-resolution images from everywhere the body. It is broadly used and regarded as one of the most accurate techniques for cancer detection and classification, because of its high-quality images on the brain tissue [7]. Tumors may have different shapes and there may not be sufficient noticeable parameters in the picture. Moreover, misdiagnosis can lead a patient to death. That is why human diagnosis is ordinarily inconsistent.

The improvement of new technologies, particularly Artificial Intelligence and Machine learning, significantly affects the clinical field, offering an error-free diagnosis. For image segmentation and classification different machine-learning methods are applied in MRI image processing. Deep learning is one kind of machine learning where layered algorithmic architecture is used to analyze data. In deep learning, model data is filtered through several layers. Each successive layer uses the result of the previous layer to make correlations and connections so that the accuracy of the model gets increased. By introducing more advanced data analysis, deep learning will make medical research more easy and reliable. The reason behind why deep learning is getting so famous in the medical research industry is because of the huge amount of data that it can process, including research data, patient outcomes, and more within a very short time with excellent accuracy.

Proper treatment of brain cancer patients depends on the proper identification of brain tumors. To diagnose the tumor by a human can be a time-consuming process and sometimes incorrect diagnosis leads a patient's life to a very critical condition. Using deep learning or machine learning, we can easily classify the tumor types, experts can make a rapid decision about treatment, and thus best treatment can be delivered to patients.

In this paper, a new system BCM-VEMT has been proposed that aims classify MRI brain images. In this system, a Convolutional Neural Network (CNN) has been built. The model has a straightforward structure. Unique and distinct features are extracted using CNN from MRI images. Then these extracted features are fed into the Machine Learning (ML) models to classify the Brain Cancer types. Finally, An ensemble of classifiers is applied that combines the ML classifier's result, and using an ensemble technique that uses the weighted average of the ML classifiers, produces a better result than any single ML classifier.

The main contributions of this paper are given below:

i) A CNN model of straightforward structure has been trained to extract deep features from MRI images to classify among three Brain Cancer types and Normal type.

ii) To apply an ensemble of classifiers for classification using five different machine learning classifiers: Random Forest (RF), Support Vector Machine (SVM), AdaBoost (AB), K-Nearest Neighbors (KNN), and Logistic Regression (LR).

iii) The system is trained using a dataset of 3787 MRI images of four types i.e. Glioma, Meningioma, Normal, and Pituitary.

iv) The Ensemble of classsifiers has achieved 97.90% accuracy, 98.59% precision, 97.90% recall, 99.16% specificity, and 98.24% F1-score for Glioma class, 98.94% accuracy, 97.89% precision, 98.94% recall, 99.30% specificity, and 98.41% F1-score for Meningioma class, 98.00% accuracy, 100.00% precision, 98.00% recall, 100.00% specificity, and 98.99% F1-score for Normal class, 98.92% accuracy, 97.87% precision, 98.92% recall, 99.30% specificity, and 98.40% F1-score for Pituitary class.

The rest of the paper is summed up as: Section 2 illustrates recent researches for the detection and classification of Brain Cancer types from MRI images by using deep neural networks and ML algorithms. Section 3 represents an all-over description of the dataset, data preprocessing, data augmentation along with the proposed models of BCM-VEMT and their structures. Section 4 shows the performance analysis of BCM-VEMT with evaluation metrics. Section 5 represents a comparative analysis of BCM-VEMT with the currently existing systems. Finally, Section 6 concludes the paper.

2 Related Work

Over the last few years, there have been so many studies proposing an automated system to detect and classify Brain Cancer. This section describes the current researches related to deep learning and the ensemble of ML classifiers to detect and classify Brain Cancers. Table 1 shows the comparative analysis of these recent researches.

Ketan Machhale et al. [8] have suggested a system that recognized the abnormal and normal brain MRI images. In their research, a Hybrid Classifier (SVM-KNN) consisting of the Support Vector Machine (SVM) and K- Nearest Neighbor (KNN) has been used for classifying 50 images. They have found 96% classification accuracy using SVM with Quadratic kernel and 98% using Hybrid classifier (SVM-KNN). Sarah Ali, Abdelaziz Ismael et al. [9] proposed Residual Networks to classify brain tumor images. They have evaluated their proposed model with a dataset that contains 3064 MRI images classified into three classes (Glioma, Meningioma, and Pituitary tumors). Their approach came out with an accuracy of 99%. Aaswad Sawant et al. [10] have used Convolutional Neural Network (CNN) with 5 layers for detecting Brain Cancer from MRI images. They have used a dataset containing 1800 MRI images out of which and 900 were non-cancerous and 900 were cancerous. Their system achieved 99% training accuracy and 98.6% validation accuracy.

Arun Kumar et al. [11] have proposed the Support Vector Machine (SVM) to predict and classify Brain Cancer. In their proposed system they have used 3 different datasets including a total of 201 images with different classes. They have achieved an accuracy of 96.51% with 94.2% specificity and 97.72% sensitivity. Zar Nawab Khan Swati et al. [12] proposed a novel feature extraction framework based on VGG-16 which is a type of deep convolutional neural network architecture. They have performed their experiment on a CE-MRI dataset that is publicly available and contains 3 types of brain tumors (Glioma, Meningioma, and Pituitary). The dataset consists of 3064 images of 233 patients across the coronal, axial, and sagittal views. This system obtained 96.13% accuracy.

Raheleh Hashemzehi et al. [13] proposed a hybrid paradigm consisting of a Convolutional Neural Network (CNN) and a Neural Autoregressive Distribution Estimation (NADE) to classify brain cancer from MRI images. They have extracted the features and automatically estimated the distribution of data. The system used a total of 3064 CE-MRI images, taken from 233 patients, that contains 1426 images of Glioma, 708 images of Meningioma, and 930 images of Pituitary brain tumor. They have shown that their proposed system achieved a classification accuracy of 95%. Hein Tun Zaw et al. [14] proposed Naïve Bayes classification to predict brain tumors using MRI images. The dataset they used was downloaded from the REMBRANDT database. This database contains a total of 114 MRI images with 90 images of tumors class and 24 of normal class. They gained an overall accuracy of 94%. R.Ezhilarasi and P.Varalakshm et al. [15] proposed AlexNet to classify different types of tumors. Moreover, to train their model Region Proposal Network (RPN) by Faster R-CNN algorithm is also used in this system. This research dataset was obtained from www.sciencesource.com and www.radiologyassistant.nl. The Dataset consists of 50 MRI images. They obtained an accuracy of above 99%.

Eman Abdel-Maksoud et al. [16] proposed an image segmentation technique that uses the K-means clustering integrated with the Fuzzy C-means algorithm. This technique can get the benefits of the K-means clustering for image segmentation within minimal computation time. They used a total of three datasets. The Digital Imaging and Communications in Medicine (DICOM) data set containing 22 images of brain tumors is the first one. The second one is the Brain Web data set containing 152 simulated brain MRI data, and the BRATS database is the last one. Their proposed system achieved an overall accuracy of 85.7%. Vida Harati et al. [17] proposed fuzzy connectedness (FC) algorithm based on a scale in which the seed point is selected automatically. They have used a dataset that was captured from 10 patients containing 124 images with a resolution of 256×256 for each patient. Their research comes out with an accuracy of 92%.

Table 1. Comparative study of current researches based on Brain Cancer detection and classification

Author	Sources of Dataset	Dataset Details	Model	Ensemble	Accuracy (%)
Ketan Machhale et al. [8]	--	50 images of Cancerous and Non-Cancerous type	SVM-KNN	Yes	98
Sarah Ali et al. [9]	Brain MRI images: [18]	Glioma (1426), Meningioma (708), and Pituitary (930)	ResNet50	No	99

Aaswad Sawant et al. [10]	Brain MRI images: [19], [20], [21], [22]	1800 images (900 Cancerous, and 900 Non-Cancerous)	CNN	No	98.6
Arun Kumar et al. [11]	DICOM dataset: [23]; Brain Web dataset: [24]	DICOM dataset-22, Brain Web dataset-44, and 135 images from expert radiologists	SVM	No	96.51
Nawab Khan Swati et al. [12]	CE-MRI dataset: [18]	Glioma (1426), Meningioma (708), and Pituitary (930)	CNN	No	96.13
. Raheleh Hashemzehi et al. [13]	CE-MRI images: [18]	Glioma (1426), Meningioma (708), and Pituitary (930)	Hybrid CNN-NADE	No	95
Hein Tun Zaw et al.[14]	Rembrandt database: [25]	114 (24 normal and 90 tumors)	Naive Bayes	No	94
P.Varalakshm et al. [15]	Brain MRI images: www.sciencesource.com, www.radiologyassistant.nl	50 brain MRI images of 320×440 pixels	Faster R-CNN	No	99
. Eman Abdel-Maksoud et al. [16]	DICOM dataset: [26]; Brain Web dataset: [27]; BRATS database: [28]	DICOM-22, Brain Web data set-152, BRATS-81	K-means clustering with Fuzzy C-means algorithm	No	85.7
Vida Harati et al. [17]	Brain MRI images: [29], [30]	124 images of three types (Glioma, Meningioma, and Astrocytoma)	fuzzy connectedness	No	92

3 Methodology

To develop the system BCM-VEMT, the dataset has been collected from various resources. The full dataset has been divided into the training set, validation set, and test set. Different preprocessing techniques have been applied to the dataset to make it appropriate for use in the proposed CNN model. The training set and validation set are used to train the proposed CNN model. The model ran up to 43 epochs which have been controlled by the EarlyStopping function to avoid overfitting and underfitting problems. After training the CNN model, the first fully connected layer (FCL) of the CNN has been chosen to extract the feature vector for each of the training images. Then these feature vectors are used to train the ML models. Finally, these ML models are ensembled using an advanced ensemble technique that combines the predictions from each ML model as per the skill and capability of ML classifiers. The performance is analyzed by using confusion matrix, accuracy, precision, recall, specificity, and F1-score. The entire architecture of BCM-VEMT is given in Fig. 1.

3.1 Description of Dataset

The dataset of BCM-VEMT consists of four classes which are Glioma, Meningioma, Normal, and Pituitary. The MRI images have been collected from three publicly available brain cancer datasets. 2356 MRI images including

1426 images of the Glioma class and 930 images of the Pituitary class have been collected from General Hospital, Tianjin Medical University, China, and Nanfang Hospital, China, during the time from 2005 to 2010. This dataset contains MRI scan images collected from anonymously chosen 233 cancer patients. Cheng et al. [31] processed the dataset for the first time to develop a classification system for brain cancer types. These MRI images of the Glioma and Pituitary classes are available at figshare [18]. 1431 MRI images including 396 images of the Normal class and 937 images of the Pituitary class have been acquired from a public dataset of Kaggle [32]. Another 98 images of the Normal class have been collected from a different public dataset of Kaggle [33]. These 98 images of the Normal class have been merged with the previously mentioned 396 images of the Normal class. The samples of these 4 classes are shown in Fig. 2. The total dataset of BCM-VEMT consists of a total of 3787 MRI images which includes 1426 images of Glioma class, 937 images of Meningioma class, 494 images of Normal class, and 930 images of Pituitary class.

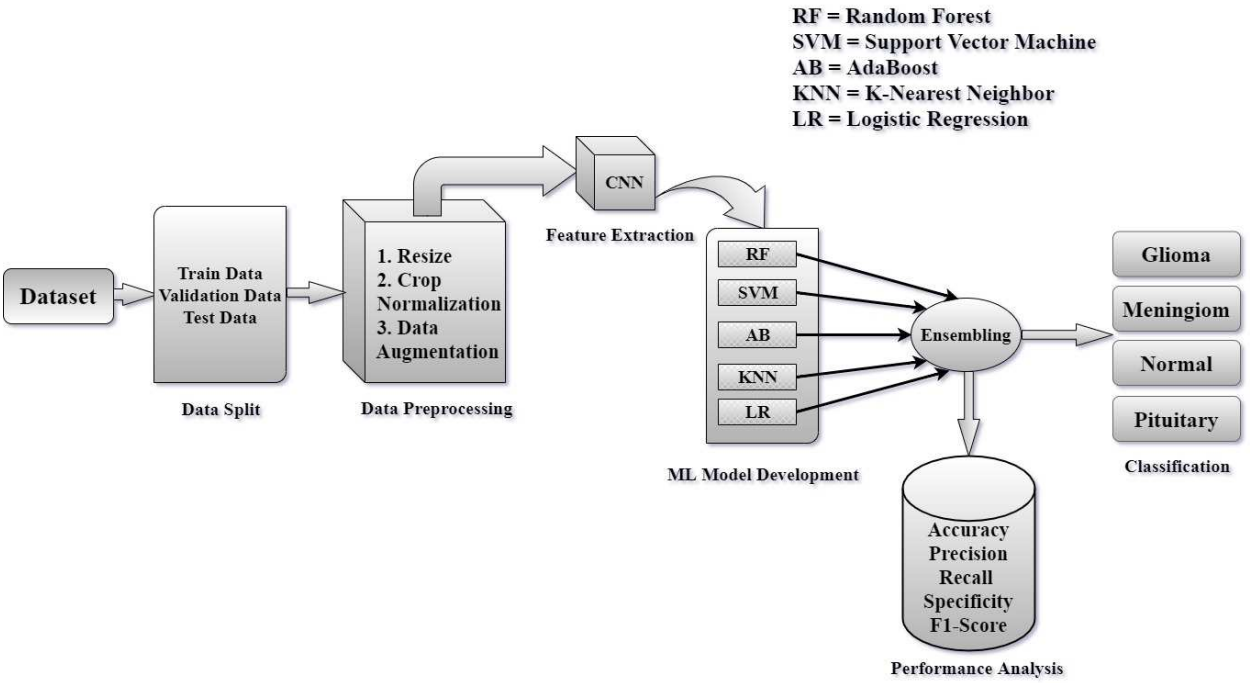


Fig. 1. Architecture of BCM-VEMT

3.2 Data Preprocessing

The MRI images of Glioma and Pituitary classes were given in MATLAB data format where each of the files stored a "struct" containing the information of the respective image. This information includes the image which was given in the form of a binary mask image, a label that differentiates the type of cancer (2 for Glioma, and 3 for Pituitary cancer), and also the image data and patient id. The images have been converted into ".jpg" format from ".mat" format and are separated into Glioma and Pituitary sets as per the label using python code. After that, the total dataset is split into training, validation, and test set with a ratio of 70:20:10 which consist of 2651 images in the training set, 756 images in the validation set, and 380 images in the test set. An overall statistical view of the BCM-VEMT dataset is shown in Table 2. All the sample images are of different sizes. To make them compatible to be used in the proposed CNN model, all the images are resized into 224×224 dimensions.

Table 2. An overall statistics of the BCM-VEMT dataset

Dataset	Glioma	Meningioma	Normal	Pituitary	Total
Training	998	656	346	651	2651
Validation	285	187	98	186	756
Test	143	94	50	93	380
Total	1426	937	494	930	3787

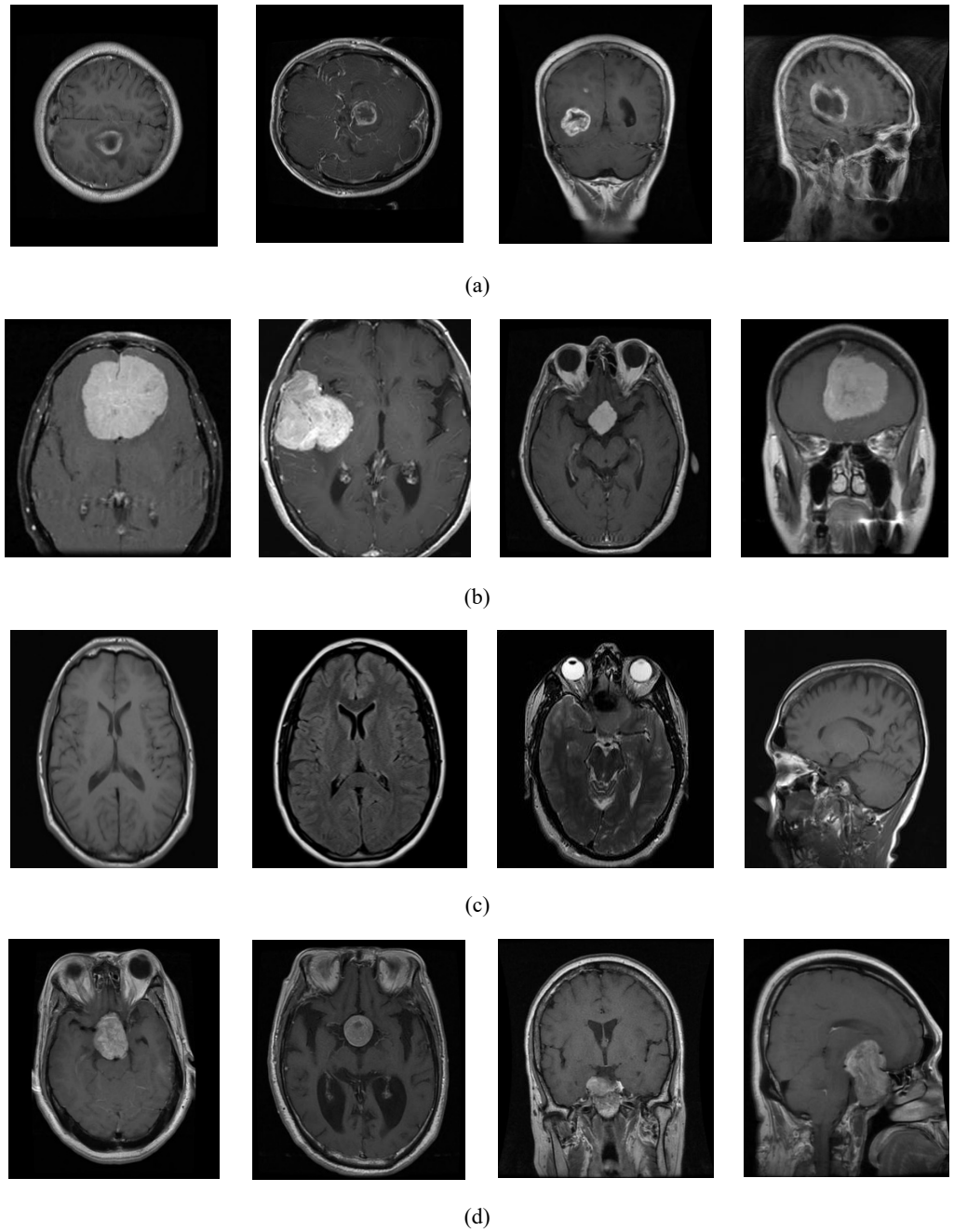
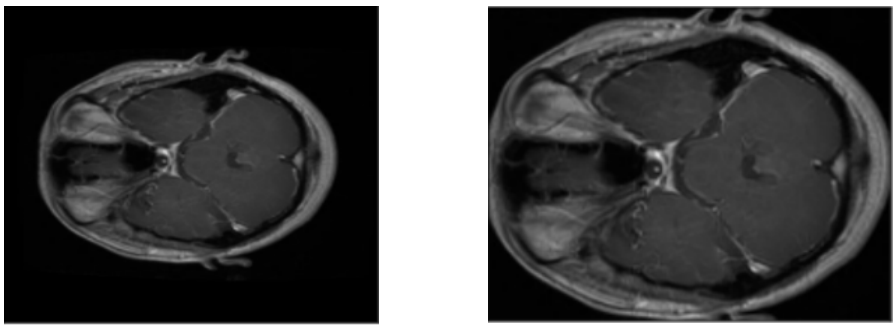


Fig. 2. Sample images of the brain cancer dataset with their class labels. (a) Glioma Samples (b) Meningioma Samples (c) Normal Samples (d) Pituitary Samples

Then, crop normalization has been performed into the dataset. This approach is performed to determine extreme points in contours. By using this crop normalization method the farthest east, west, north, and south (x, y)-coordinates along the largest contour from a given image can be determined. This technique can be applied to both rotated bounding boxes and raw contours. In BCM-VEMT, this technique is used to crop out only the portion of the image containing the brain using a Computer Vision (CV) technique. By this, the extra backgrounds of the images have been removed so that the system can focus more on the actual brain image features. Fig. 3 illustrates the aforesaid procedure.



Input Output

Fig. 3. Finding extreme points in contours and clipping them

3.3 Data Augmentation

Deep learning techniques always work better on bigger datasets. One of the ways of overcoming the lacking of a sufficient amount of training data is doing data augmentation. The term data augmentation refers to the process of expanding the training set by applying different types of manipulation techniques on the existing training set preserving the label. By applying these kinds of augmentation techniques, the size of the training set can be increased in runtime which helps to achieve better performance. In this system, data augmentation has been applied during training to accrue a better performance. As a part of data augmentation vertical and horizontal flips, rotation, shifting in both vertical and horizontal axis, shearing, and brightness manipulation have been applied to the training set. These augmentation techniques help the system to achieve better performance.

3.4 Proposed CNN Architecture

3.4.1 VGG-16

VGG-16 is a sort of Convolutional Neural Network architecture that consists of 16 layers. The uniqueness of the VGG-16 model architecture is that instead of having more parameters, it focuses on a convolution layer of 3x3 kernel size. This model's minimum input image size expectation is 224×224 pixels with 3 channels. The benefit of having small size kernels is that the overfitting problem can be avoided. Activation functions are used in the neural network to decide whether a neuron should be activated or not by calculating the weighted sum of input. A neural network contains neurons that work in coordination with weight, 10 bias, and respective activation functions. The updating of weights and biases of the neurons are done depending on the error at the output. The activation function brings non-linearity to the input of neural nets and allows them to learn and perform complex tasks.

3.4.2 Transfer Learning

Transfer learning refers to the procedure of using a model in a classification problem that was previously trained for a different classification problem. Transfer learning is commonly used to overcome the problem of insufficient data and to reduce training time. In deep learning, many pre-trained models on “Image Data” are available which can be used to train another similar classification problem with or without adding some additional layers into that pre-trained model to fit it into the new model more efficiently as per the system demand.

To train BCM-VEMT, VGG-16 has been used as Convolutional Neural Network (CNN). A pre-trained VGG-16 model with some additional layers has been used to train the system. Firstly, the pre-trained VGG-16 model has been loaded as the base model. Then a sequential model is created and the base model is added to it. A flatten layer is used so that it can convert the pooled feature map into a single column which will be passed into the fully connected layer. Two dropout layers with a threshold of 0.5 each have been used to prevent the model from overfitting. A fully connected layer having 128 neurons is added to the model to extract the features from MRI images. Another fully connected layer having neuron size as per the number of classes with a “softmax” activation function is added as the output layer. The architecture of the proposed CNN model is shown in Fig. 4.

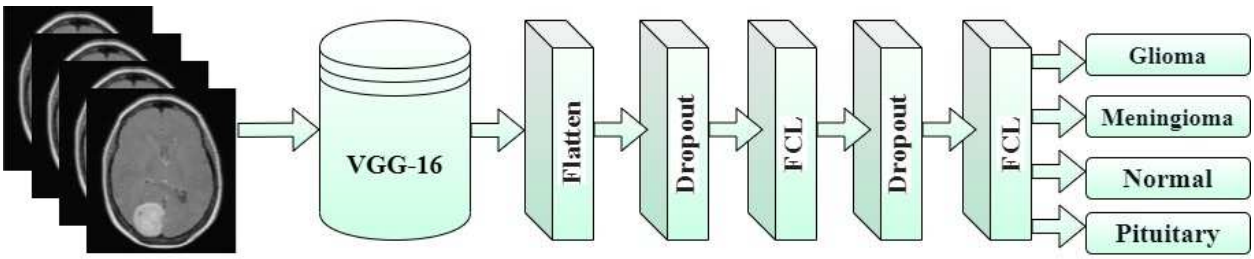


Fig. 4. The Proposed CNN Architecture

In BCM-VEMT minimum training loss is monitored with a patience value of 10. The model optimizer “RMSprop” has been used in BCM-VEMT with a learning rate of 0.0001. The summary of the proposed CNN model is shown in Table 3.

Table 3. Model summary of the proposed CNN model of BCM-VEMT

Number of Layers	Layer (Type)	Output Shape	Parameter
1	vgg16 (Functional)	(None, 7, 7, 512)	14714688
2	flatten (Flatten)	(None, 25088)	0
3	dropout (Dropout)	(None, 25088)	0
4	dense (Dense)	(None, 128)	3211392
5	dropout_1 (Dropout)	(None, 128)	0
6	dense_1 (Dense)	(None, 4)	516

3.5 ML and Ensemble Technique

3.5.1 Random Forest Classifier

Random Forest Classifier is a type of supervised machine learning algorithm. It consists of a large number of individual decision trees that operate as an ensemble and are used for both regression and classification problems. It reduces the overfitting problem by averaging the result.

3.5.2 Support Vector Machine

Support Vector Machine finds a hyperplane in N-dimensional space (N represents the number of features) which uniquely classifies the data points.

3.5.3 AdaBoost Classifier

AdaBoost, a short form of Adapting Boosting uses the ensemble method in Machine Learning. It attempts to build a strong classifier from the number of strong classifiers to increase the accuracy of the classifiers. The classifier ought to be trained interactively on different weighted training examples. In every iteration, it attempts to give an excellent fit to these examples by limiting training errors.

3.5.4 K-Nearest Neighbors

K-Nearest Neighbors, commonly known as KNN is a non-parametric and lazy learning algorithm. For the appearance of new data, KNN can classify them into a well-suited category.

3.5.5 Logistic Regression

Logistic Regression is a statistical method used to classify a categorical target variable. It uses a logistic function to estimate the probability of a target variable. The logistic curves relate the independent variable X, to the rolling mean of the DV, $P(\overline{Y})$ .The formula can be written as (1) or (2).

$$P = \frac{e^{a+bX}}{1+e^{a+bX}} \tag{1}$$

Or

$$P = \frac{1}{1+e^{-(a+bX)}} \tag{2}$$

Where the probability P is the base of the natural logarithm; and a, b are the parameters of the model.

3.5.6 Weighted Average Ensemble

Weighted Average Ensemble is an advanced ensemble technique that combines the prediction of multiple models as per the capability and skill of each model. While combining the models, it assigns a weight to each model which is proportional to its capability of correctly classifying the classification problem. A model that performs better is given a bigger weight and thus can play a vital role while making the decision. So, it can produce a better decision combining the predictions of all the used ML classifiers.

In BCM-VENT, the feature vectors are extracted from the Fully Connected Layer (FCL) of the CNN. From Table 3, it can be seen that the first FCL is “dense” which has 128 neurons. So, the feature vector will have a dimension of (1×128). BCM-VEMT has a training set of size 2651 and deep features are extracted from each of the training images. That means each image of this training set will have 128 features individually. So, to train the ML classifiers the input set dimension will be (2651×128). This input data is passed into five ML Classifiers: Random Forest (RF), Support Vector Machine (SVM), AdaBoost (AB), K-Nearest Neighbors (KNN), and Logistic Regression (LR). Each of these ML classifiers individually shows better performance than the CNN model performance. Then ensemble classifier is developed combining all the five ML classifiers. In BCM-VEMT, the

'accuracy' result of each ML classifier has been chosen as 'weight' rather than assuming each model is equally effective. BCM-VEMT assigns the 'accuracy' result to the weight parameter for each model, combines the prediction of ML models with weights associated with them, and takes a final decision. This ensemble technique performed better than any individual ML model's performance and achieved a higher accuracy. The process of training the ML classifiers and the ensemble technique is shown in Fig. 5.

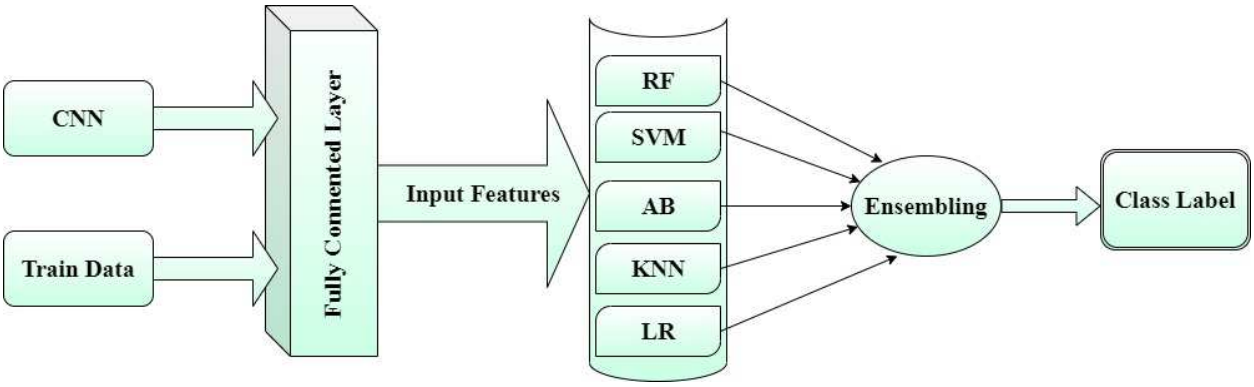


Fig. 5. Training process of the ML classifiers and ensemble of classifiers

3.6 Matrices for Performance Evaluation

To evaluate the performance of BCM-VEMT five matrices: accuracy, precision, recall, specificity, and F1-score have been used. To know about these matrices we need to know about some terms first:

True Positive (TP): Prediction and observation both are positive

True Negative (TN): Prediction and observation both are negative.

False Positive (FP): Prediction is positive though the observation is negative.

False Negative (FN): Prediction is negative though the observation is positive.

These four terms are used to compute the performance matrices of BCM-VEMT. The formulas of deriving these matrices from these four terms are shown in (3), (4), (5), (6), and (7).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

4 Experimental Analysis

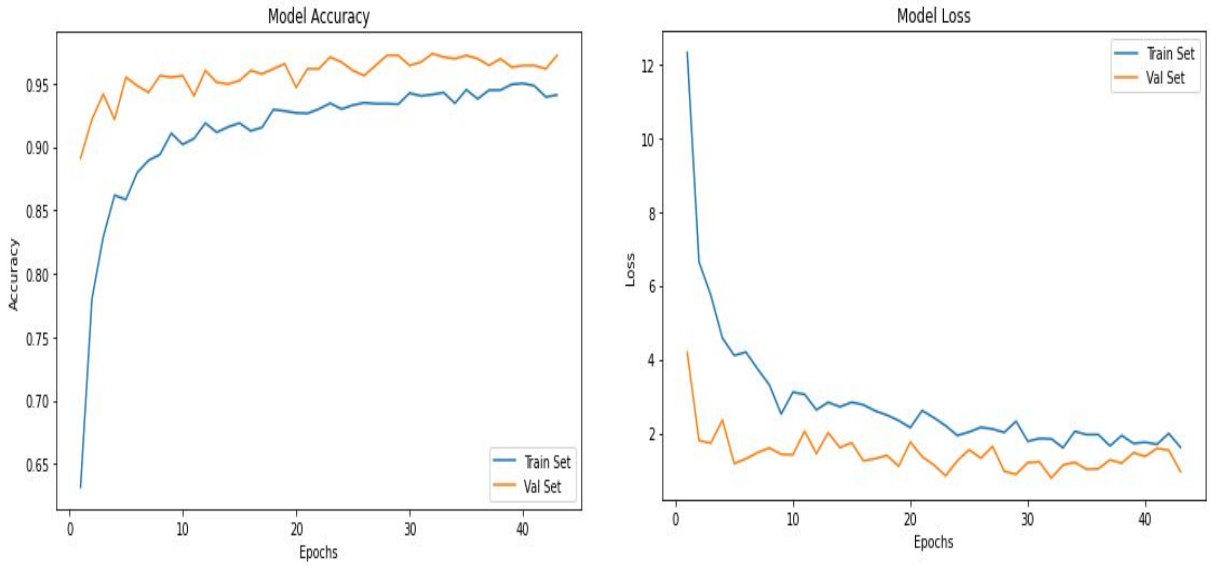
In BCM-VEMT, the CNN model was trained with the training and validation set and then the features were extracted from it to classify Brain Cancer by an ensemble of five ML models. The CNN model ran up to 43 epochs. To control the number of epochs, EarlyStopping is used in the model which helps to overcome overfitting and underfitting problems. EarlyStopping is a regularization method that monitors a defined parameter with a pre-define patience value to automatically halt the epoch running when the system stops improving. Finally, the performance of BCM-VEMT is analyzed with confusion matrices, accuracy, precision, recall, specificity, and f1-score.

4.1 Experimental Setup

The experiment was run on Google Collaboratory which is a cloud service of Google based on Jupyter Notebook. Python is used for its vast library facility and simplicity. Google Drive was used to import the dataset. Google Collaboratory provided 68.35 GB DISK and 12.69 GB RAM with access to a free-of-charge stable GPU.

4.2 Result Analysis

The training and validation accuracy of the CNN model used in BCM-VEMT is shown in Fig. 6. Training accuracy is 95.00% and loss is 1.61. Validation accuracy is 97.34% and loss is 0.79.



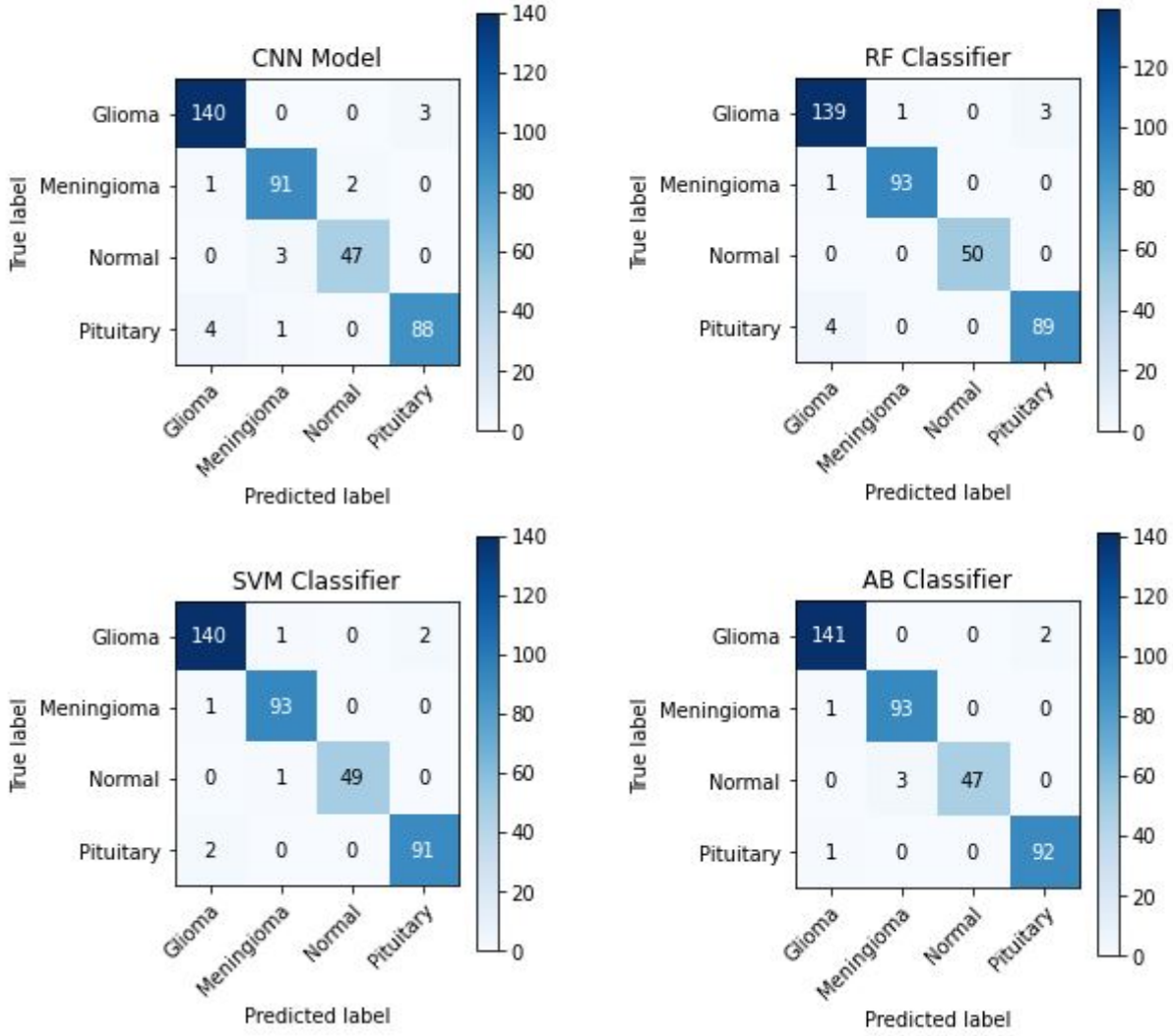
Training and Validation Accuracy

Training and Validation Loss

Fig. 6. Performance analysis of the CNN model used in BCM-VEMT

Fig. 7 shows the confusion matrix of the CNN model along with the five ML classifier and the ensemble of classifiers. The test set has a total of 380 images including 143 Glioma cases, 94 Meningioma cases, 50 Normal cases, and 93 Pituitary cases. In confusion matrices, true labels are placed along the row, and predicted labels are placed along the column. The CNN model correctly classifies 140 Glioma cases, 91 Meningioma cases, 47 Normal cases, and 88 Pituitary cases, and on the other hand misclassifies 3 cases of each Glioma, Meningioma and Normal

class along with 5 Pituitary cases. The RF classifier truly classifies 139 Glioma cases, 93 Meningioma cases, all 50 Normal cases, and 89 Pituitary cases and in addition misjudges 4 Glioma cases, single Meningioma case, and 4 Pituitary cases. The SVM classifier perfectly classifies 140 Glioma cases, 93 Meningioma cases, 49 Normal cases, and 91 Pituitary cases and in contrariwise fails to classify 3 Glioma cases, single Meningioma and Normal case, and 2 Pituitary cases. The AB classifier succeeds to classify 141 Glioma cases, 93 Meningioma cases, 47 Normal cases, and 92 Pituitary cases perfectly whereas shows a failure in classifying 2 Glioma cases, single Meningioma case, 3 Normal cases, and a single Pituitary case.



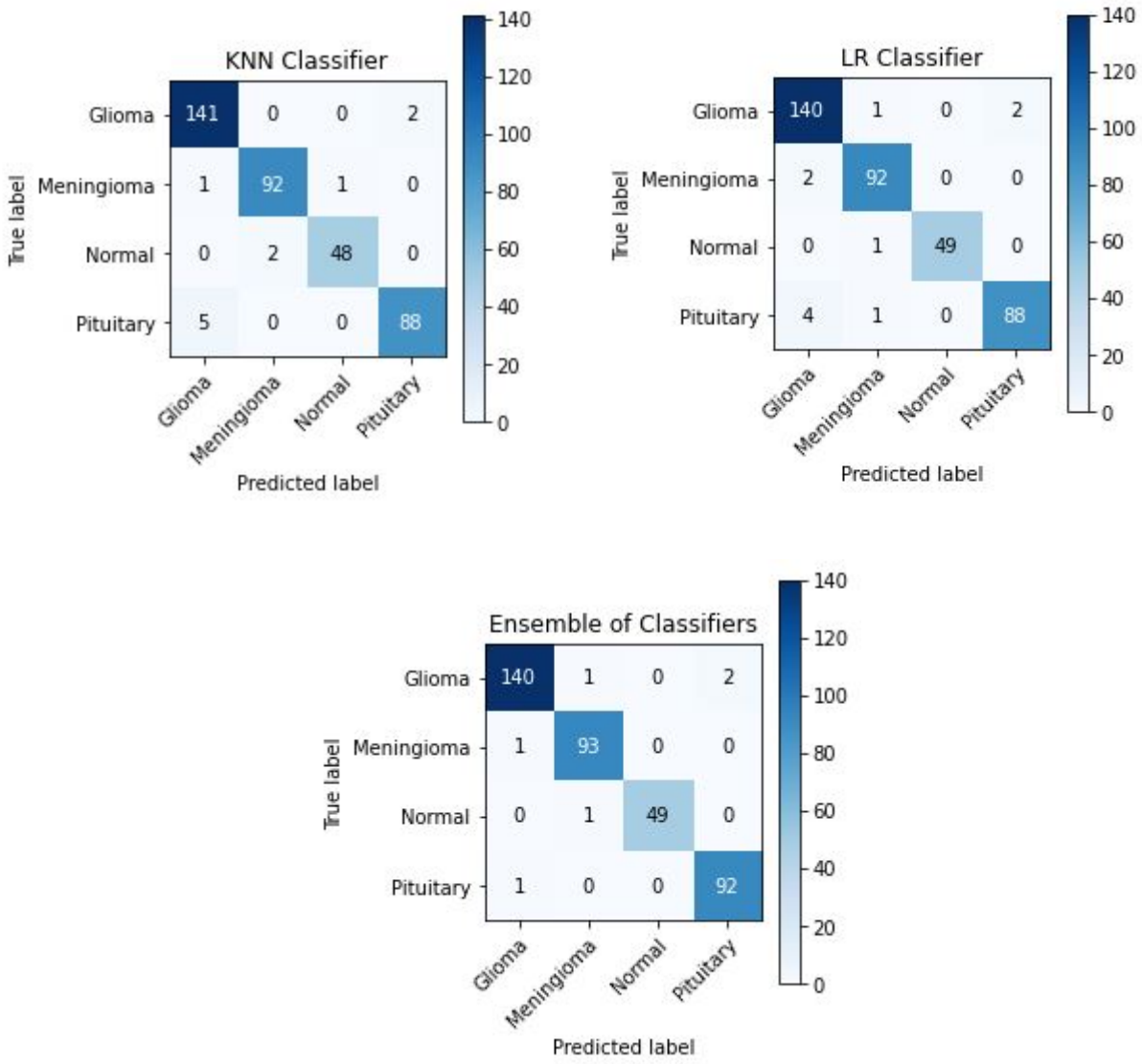


Fig. 7. Confusion matrix representation of the classifiers used in BCM-VEMT

The KNN classifier correctly classifies 141 Glioma cases, 92 Meningioma cases, 48 Normal cases, and 88 Pituitary cases but missed 2 of each Glioma, Meningioma, and Normal case along with 5 Pituitary cases to classify those in their true class. The LR classifier perfectly classifies 140 Glioma cases, 92 Meningioma cases, 49 Normal cases, and 88 Pituitary cases and in contrariwise fails to classify 3 Glioma cases, 2 Meningioma cases, single Normal case, and 5 Pituitary cases correctly. Finally, The Ensemble of classifiers achieved the best performance by correctly classifying 140 Glioma cases, 93 Meningioma cases, 49 Normal cases, and 92 Pituitary cases.

The evaluation matrices of BCM-VEMT: accuracy, precision, recall, specificity, and f1-score of classifiers for each class (Glioma, Meningioma, Normal, and Pituitary) are shown in Table 4. CNN model achieved 97.90% accuracy, 96.55% precision, 97.90% recall, 97.89% specificity, and 97.22% F1-score for Glioma class, 96.80% accuracy, 95.78% precision, 96.80% recall, 98.60% specificity, and 96.29% F1-score for Meningioma class, 94.00% accuracy, 95.92% precision, 94.00% recall, 99.39% specificity, and 94.95% F1-score for Normal class, 94.62% accuracy, 96.70% precision, 94.62% recall, 98.95% specificity, and 95.65% F1-score for Pituitary class. RF model

achieved 97.20% accuracy, 96.52% precision, 97.20% recall, 97.89% specificity, and 96.86% F1-score with Glioma cases, 98.94% accuracy, 98.94% precision, 98.94% recall, 99.65% specificity, and 98.94% F1-score with Meningioma cases, 100.00% accuracy, 100.00% precision, 100.00% recall, 100% specificity, and 100.00% F1-score with Normal cases, 95.70% accuracy, 96.74% precision, 95.70% recall, 98.95% specificity, and 96.22% F1-score with Pituitary cases. SVM model gained 97.90% accuracy, 97.90% precision, 97.90% recall, 98.73% specificity, and 99.90% F1-score for Glioma class, 98.94% accuracy, 97.89% precision, 98.94% recall, 99.30% specificity, and 98.41% F1-score for Meningioma class, 98.00% accuracy, 100.00% precision, 98.00% recall, 100.00% specificity, and 98.99% F1-score for Normal class, 97.85% accuracy, 97.85% precision, 97.85% recall, 99.30% specificity, and 97.85% F1-score for Pituitary class. AB model achieved 98.60% accuracy, 98.60% precision, 98.60% recall, 99.15% specificity, and 98.60% F1-score with Glioma cases, 98.94% accuracy, 96.87% precision, 98.93% recall, 98.95% specificity, and 97.89% F1-score with Meningioma cases, 94.00% accuracy, 100.00% precision, 94.00% recall, 100% specificity, and 96.91% F1-score with Normal cases, 98.92% accuracy, 97.87% precision, 98.92% recall, 99.30% specificity, and 98.40% F1-score with Pituitary cases. KNN model achieved 98.60% accuracy, 95.92% precision, 98.60% recall, 97.47% specificity, and 97.24% F1-score for Glioma class, 97.87% accuracy, 97.87% precision, 97.87% recall, 99.30% specificity, and 97.87% F1-score for Meningioma class, 96.00% accuracy, 97.96% precision, 96.00% recall, 99.70% specificity, and 96.97% F1-score for Normal class, 94.62% accuracy, 97.78% precision, 94.62% recall, 99.30% specificity, and 96.17% F1-score for Pituitary class. LR model achieved 97.90% accuracy, 95.89% precision, 97.90% recall, 97.47% specificity, and 96.89% F1-score with Glioma cases, 97.87% accuracy, 96.84% precision, 97.87% recall, 98.95% specificity, and 97.35% F1-score with Meningioma cases, 98.00% accuracy, 100.00% precision, 98.00% recall, 100.00% specificity, and 98.99% F1-score with Normal cases, 94.62% accuracy, 97.78% precision, 94.62% recall, 99.30% specificity, and 96.17% F1-score with Pituitary cases.

Table 4. Performance evaluation of the classifiers used in BCM-VEMT based on each class label

Model	Class	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 - Score (%)
CNN	Glioma	97.90	96.55	97.90	97.89	97.22
	Meningioma	96.80	95.78	96.80	98.60	96.29
	Normal	94.00	95.92	94.00	99.39	94.95
	Pituitary	94.62	96.70	94.62	98.95	95.65
RF	Glioma	97.20	96.52	97.20	97.89	96.86
	Meningioma	98.94	98.94	98.94	99.65	98.94
	Normal	100.00	100.00	100.00	100.00	100.00
	Pituitary	95.70	96.74	95.70	98.95	96.22
SVM	Glioma	97.90	97.90	97.90	98.73	97.90
	Meningioma	98.94	97.89	98.94	99.30	98.41
	Normal	98.00	100.00	98.00	100.00	98.99
	Pituitary	97.85	97.85	97.85	99.30	97.85
AB	Glioma	98.60	98.60	98.60	99.15	98.60
	Meningioma	98.94	96.87	98.93	98.95	97.89
	Normal	94.00	100.00	94.00	100.00	96.91
	Pituitary	98.92	97.87	98.92	99.30	98.40
KNN	Glioma	98.60	95.92	98.60	97.47	97.24
	Meningioma	97.87	97.87	97.87	99.30	97.87
	Normal	96.00	97.96	96.00	99.70	96.97

	Pituitary	94.62	97.78	94.62	99.30	96.17
LR	Glioma	97.90	95.89	97.90	97.47	96.89
	Meningioma	97.87	96.84	97.87	98.95	97.35
	Normal	98.00	100.00	98.00	100.00	98.99
	Pituitary	94.62	97.78	94.62	99.30	96.17
Ensemble	Glioma	97.90	98.59	97.90	99.16	98.24
	Meningioma	98.94	97.89	98.94	99.30	98.41
	Normal	98.00	100.00	98.00	100.00	98.99
	Pituitary	98.92	97.87	98.92	99.30	98.40

Finally the Ensemble of classifiers achieved 97.90% accuracy, 98.59% precision, 97.90% recall, 99.16% specificity, and 98.24% F1-score for Glioma class, 98.94% accuracy, 97.89% precision, 98.94% recall, 99.30% specificity, and 98.41% F1-score for Meningioma class, 98.00% accuracy, 100.00% precision, 98.00% recall, 100% specificity, and 98.99% F1-score for Normal class, 98.92% accuracy, 97.87% precision, 98.92% recall, 99.30% specificity and 98.40% F1-score for Pituitary class.

The system's overall accuracy is shown in Table 5. Initially, the CNN Model achieved 96.32% accuracy. Then each ML model gained individually higher accuracy than the CNN model including 97.63% accuracy for the RF model, 98.15% accuracy for the SVM model, 98.15% accuracy for the AB model, 97.11% accuracy for the KNN model, and 97.11% accuracy for LR model. Among all five ML models, SVM and AB models achieved the highest accuracy. That means these two models played the most vital role in Ensembling. Finally, the Ensemble classifier applying "Weighted Average Ensemble" on these ML models achieved the final accuracy of 98.42%.

Table 5. Performance evaluation of the classifiers used in BCM-VEMT

Model	CNN	RF	SVM	AB	KNN	LR	Ensemble
Accuracy (%)	96.32	97.63	98.15	98.15	97.11	97.11	98.42

The ROC curve of each class for the Ensemble of classifiers is shown in Fig. 7. In the curve, False Positive Rate (FPR), and True Positive Rate (TPR) is plotted on the x-axis and y-axis respectively. A higher value of area under the curve (AUC) of ROC reflects a better efficiency of the model in diagnosing the true classes of the classification problem. The higher the AUC value is the better the model is. Fig. 7 shows the value of AUC is 0.9867 for the Glioma class, 0.9877 for the Meningioma class, 0.9985 for Normal Class, and 0.9876 for the Pituitary class. These values represent that BCM-VEMT has a higher efficiency in detecting each class at its true label. Moreover, the Normal Class has an AUC of 0.9985 which is a very positive sign for medical diagnosis problems. So, the model can have a good contribution in classifying these types of Brain Cancers.

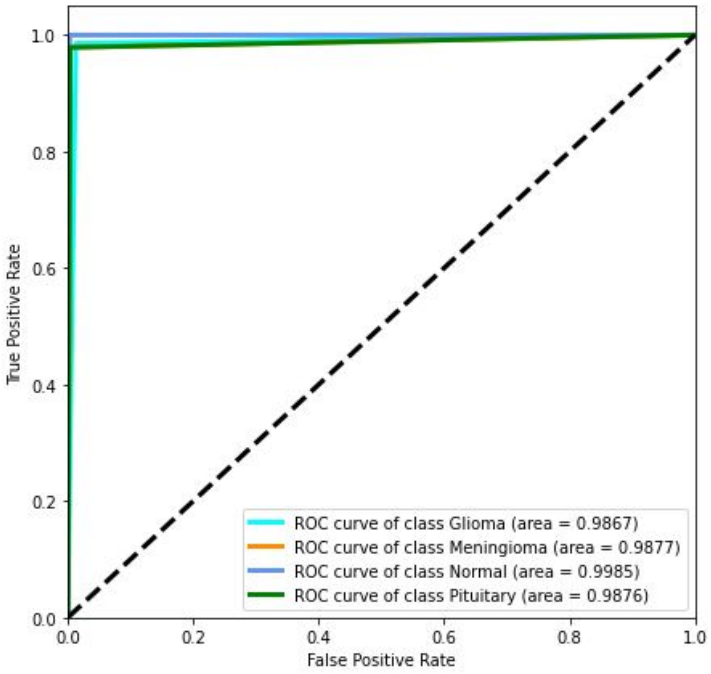


Fig. 7. ROC Curve of each class for Ensemble of Classifiers

5 Discussion

Table 6 shows the comparative performance evaluation of BCM-VEMT with the recent studies about Brain Cancer. The system proposed in [16], [17], [14], [13], gained the accuracy of 85.7%, 92%, 94%, and 95% respectively. Then the systems proposed in [12], [11], and [8] achieved the accuracy of 96.13%, 96.51%, and 98% respectively. All these systems are for two or three classes. BCM-VEMT outperforms all these systems with an accuracy of 98.42% with four different classes.

Table 6. Performance comparison of BMC-VEMT with recent studies

Author	Sources of Dataset	Dataset Details	Model	Ensemble	Accuracy (%)
Ketan Machhale et al. [8]	--	50 images of Cancerous and Non-Cancerous type	SVM-KNN	Yes	98
Sarah Ali et al. [9]	Brain MRI images: [18]	Glioma (1426), Meningioma (708), and Pituitary (930)	ResNet50	No	99
Aaswad Sawant et al. [10]	Brain MRI images: [19], [20], [21], [22]	1800 images (900 Cancerous, and 900 Non-Cancerous)	CNN	No	98.6

Arun Kumar et el. [11]	DICOM dataset: [23]; Brain Web dataset: [24]	DICOM dataset-22, Brain Web dataset-44, and 135 images from expert radiologists	SVM	No	96.51
Nawab Khan Swati et al. [12]	CE-MRI dataset: [18]	Glioma (1426), Meningioma (708), and Pituitary (930)	CNN	No	96.13
. Raheleh Hashemzehi et al. [13]	CE-MRI images: [18]	Glioma (1426), Meningioma (708), and Pituitary (930)	Hybrid CNN-NADE	No	95
Hein Tun Zaw et al.[14]	Rembrandt database: [25]	114 (24 normal and 90 tumors)	Naive Bayes	No	94
P.Varalakshm et al. [15]	Brain MRI images: www.sciencesource.com, www.radiologyassistant.nl	50 brain MRI images of 320×440 pixels	Faster R-CNN	No	99
. Eman Abdel-Maksoud et al. [16]	DICOM data set: [26]; Brain Web data set: [27]; BRATS database: [28]	DICOM-22, Brain Web data set-152, BRATS-81	K-means clustering with Fuzzy C-means algorithm	No	85.7
Vida Harati et al. [17]	Brain MRI images: [29], [30]	124 images of three types (Glioma, Meningioma, and Astrocytoma)	fuzzy connectedness	No	92
BCM-VEMT	**Brain MRI Images: [18], [32],[33]**	**3787 (1426 Glioma , 937 Meningioma, 494 Normal, and 930 Pituitary)**	**CNN + Ensemble of ML Models**	**Yes**	**98.42**

The system proposed in [10], [9], and [15] gained outstanding performance with an accuracy of 98.6%, 99%, and 99% which are higher than BMC-VEMT. But in [10], there are only two classes as cancerous and non-cancerous and in [9] there are three classes (Glioma, Meningioma, and Pituitary tumor). In [15] there are four classes (benign, malignant, glial, and astrocytoma), but the dataset they used contains only 50 images. On the other hand, our proposed system has 4 classes including Glioma, Meningioma, Normal, and Pituitary with 3787images and achieved an accuracy of 98.42%. Comparing with those systems our BCM-VEMT gives a better performance that will help the experts to classify Brain Cancers more reliably and more accurately. Besides having this excellent performance the system BCM-VEMT still has some limitations such that though it can predict the majority of the samples in its actual class it still misclassifies some samples. To make the system more accurate and efficient, in the future some more advanced models and classifiers will have experimented. Moreover, there is a plan to train the system with a bigger dataset to improve its performance.

6 Conclusion

The number of infected people with brain cancer is growing high day by day all over the world. In this paper, an enhanced approach has been proposed that can classify brain cancer types using deep learning and an ensemble of machine learning techniques on MRI images. Various kinds of data pre-processing and data augmentation techniques have been performed to enhance the quality of the images and to increase the overall accuracy. The

system was trained on a dataset of a total of 3787 images of 4 classes. The performance of BCM-VEMT was measured under several matrices such as Accuracy, Precision, Recall, Specificity, and F1-Score. BCM-VEMT has achieved 97.90% accuracy, 98.59% precision, 97.90% recall, 99.16% specificity, and 98.24% F1-score for Glioma class, 98.94% accuracy, 97.89% precision, 98.94% recall, 99.30% specificity, and 98.41% F1-score for Meningioma class, 98.00% accuracy, 100.00% precision, 98.00% recall, 100.00% specificity, and 98.99% F1-score for Normal class, 98.92% accuracy, 97.87% precision, 98.92% recall, 99.30% specificity and 98.40% F1-score for Pituitary class with an overall accuracy of 98.42%. From all this statistical analysis it is clear that BCM-VEMT is showing an excellent performance with high accuracy and efficiency in classifying Brain Cancer types. So, it can play a vital role as a helping hand of doctors to detect and classify Brain Cancer at an early stage.

Acknowledgments: We are grateful to all data hosting providers for their support, and storage capacity to deliver datasets for our research.

References

[1] Montreal Children's Hospital. 2021. *True or False? Not all tumors are cancerous.* [online] Available at: <https://www.thechildren.com/health-info/conditions-and-illnesses/true-or-false-not-all-tumors-are-cancerous> [Accessed 26 July 2021].

[2] Medicalnewstoday.com. 2021. *Tumors: Benign, premalignant, and malignant*. [online] Available at: <https://www.medicalnewstoday.com/articles/249141> [Accessed 26 July 2021].

[3] Nan Zhang, Su Ruan, Stephane Lebonvallet, Qingmin Liao and Yuemin Zhu. Kernel Feature Selection to Fuse Multi-spectral MRI Images for Brain Tumor Segmentation. Computer Vision and Image Understanding, 2011, 115(2):256-269.

[4] National Brain Tumor Society. 2021. *Quick Brain Tumor Facts | National Brain Tumor Society*. [online] Available at: <https://braintumor.org/brain-tumor-information/brain-tumor facts/#:~:text=Today%2C%20an%20estimated%20700%2C000%20people,patient%20and%20their%20loved%20ones.> [Accessed 4 July 2021].

[5] Cancer Treatment Centers of America. 2021. *Types of Brain Cancer: Common, Rare and More Varieties*. [online] Available at: <https://www.cancercenter.com/cancer-types/brain-cancer/types> [Accessed 26 July 2021].

[6] Swati, Z., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S. and Lu, J., 2019. Content-Based Brain Tumor Retrieval for MR Images Using Transfer Learning. *IEEE Access*, 7, pp.17809-17822.

[7] Kumar, S., Dabas, C. and Godara, S., 2017. Classification of Brain MRI Tumor Images: A Hybrid Approach. *Procedia Computer Science*, 122, pp.510-517.

[8] K. Machhale, H. B. Nandpuru, V. Kapur, and L. Kosta, "MRI brain cancer classification using hybrid classifier (SVM-KNN)," *2015 Int. Conf. Ind. Instrum. Control. ICIC 2015*, no. Icic, pp. 60–65, 2015, doi: 10.1109/IIC.2015.7150592.

[9] S. A. Abdelaziz Ismael, A. Mohammed, and H. Hefny, "An enhanced deep learning approach for brain cancer MRI images classification using residual networks," *Artif. Intell. Med.*, vol. 102, p. 101779, 2020, doi: 10.1016/j.artmed.2019.101779.

[10] A. Sawant, M. Bhandari, R. Yadav, R. Yele, and S. Bendale, "Brain Cancer Detection From Mri: a Machine Learning Approach (Tensorflow)," *Int. Res. J. Eng. Technol.*, vol. 9001, p. 2089, 2008, [Online]. Available: www.irjet.net.

[11] N. B. Bahadure, A. K. Ray, and H. P. Thethi, "Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM," *Int. J. Biomed. Imaging*, vol. 2017, 2017, doi: 10.1155/2017/9749108.

[12] Z. N. K. Swati *et al.*, "Content-Based Brain Tumor Retrieval for MR Images Using Transfer Learning," *IEEE Access*, vol. 7, no. c, pp. 17809–17822, 2019, doi: 10.1109/ACCESS.2019.2892455.

[13] R. Hashemzehi, S. J. S. Mahdavi, M. Kheirabadi, and S. R. Kamel, "Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE," *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 1225–1232, 2020, doi: 10.1016/j.bbe.2020.06.001.

[14] H. T. Zaw, N. Maneerat, and K. Y. Win, "Brain tumor detection based on Naïve Bayes classification," *Proceeding - 5th Int. Conf. Eng. Appl. Sci. Technol. ICEAST 2019*, pp. 1–4, 2019, doi: 10.1109/ICEAST.2019.8802562.

[15] R. Ezhilarasi and P. Varalakshmi, "Tumor detection in the brain using faster R-CNN," *Proc. Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud), I-SMAC 2018*, pp. 388–392, 2019, doi: 10.1109/I-SMAC.2018.8653705.

[16] E. Abdel-Maksoud, M. Elmogy, and R. Al-Awadi, "Brain tumor segmentation based on a hybrid clustering technique," *Egypt. Informatics J.*, vol. 16, no. 1, pp. 71–81, 2015, doi: 10.1016/j.eij.2015.01.003.

[17] V. Harati, R. Khayati, and A. Farzan, “Fully automated tumor segmentation based on improved fuzzy connectedness algorithm in brain MR images,” *Comput. Biol. Med.*, vol. 41, no. 7, pp. 483–492, 2011, doi: 10.1016/j.compbiomed.2011.04.010.

[18] figshare. 2021. *brain tumor dataset*. [online] Available at: <https://figshare.com/articles/dataset/brain_tumor_dataset/1512427> [Accessed 4 July 2021].

[19] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Γ‡, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694.

[20] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117.

[21] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. (2017) Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q.

[22] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Michel Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, Christos Davatzikos. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 doi: 10.1038/sdata.2017.117 (2017).

[23] Osirix-viewer.com. 2021. *OsiriX DICOM Viewer | The world famous medical imaging viewer*. [online] Available at: <http://www.osirix-viewer.com/.> [Accessed 26 July 2021].

[24] 2021. [online] Available at: <http://brainweb.bic.mni .mcgill.ca/cgi/brainweb1.> [Accessed 26 July 2021].

[25] 2021. [online] Available at: <https://wiki.cancerimagingarchive.net/display/Public/REMBRA> [Accessed 4 July 2021].

[26] Brainweb.bic.mni.mcgill.ca. 2021. *BrainWeb: Simulated Brain Database*. [online] Available at: <https://brainweb.bic.mni.mcgill.ca/brainweb/> [Accessed 4 July 2021].

[27] Osirix-viewer.com. 2021. *OsiriX DICOM Viewer | DICOM Image Library*. [online] Available at: <http://www.osirix-viewer.com/resources/dicom-image-library/> [Accessed 4 July 2021].

[28] Www2.imm.dtu.dk. 2021. *MICCAI BRATS 2012*. [online] Available at: <http://www2.imm.dtu.dk/projects/BRATS2012/data.html> [Accessed 4 July 2021].

[29] M. Kaus, S.K. Warfield, A. Nabavi, P.M. Black, F.A. Jolesz, R. Kikinis, Automated segmentation of MRI of brain tumors, Radiology 218 (2) (2001) 586–591, /http://www.spl.harvard.edu/Special:PubDB_View?dspaceid=169S.

[30] S.K. Warfield, M. Kaus, F.A. Jolesz, R. Kikinis, Adaptive, template moderated, spatially varying statistical classification, Med. Image Anal. 4 (1) (2000) 43–55, /http://www.spl.harvard.edu/pages/SoftwareS.

[31] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, and Z. Yun, “Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition,” pp. 1–13, 2015, doi: 10.1371/journal.pone.0140381.

[32] Kaggle.com. 2021. *Brain Tumor Classification (MRI)*. [online] Available at: <https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri> [Accessed 4 July 2021].

[33] Kaggle.com. 2021. *Brain MRI Images for Brain Tumor Detection*. [online] Available at: <https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection> [Accessed 4 July 2021].