VGG-SCNet: A VGG Net-Based Deep Learning Framework for Brain Tumor Detection on MRI Images

MOHAMMAD SHAHJAHAN MAJIB[1], MD. MAHBUBUR RAHMAN[1], (Member, IEEE), T. M. SHAHRIAR SAZZAD[1], NAFIZ IMTIAZ KHAN[1], AND SAMRAT KUMAR DEY[2]

[1]Department of Computer Science and Engineering, Military Institute of Science and Technology, Dhaka 1216, Bangladesh
[2]Department of Computer Science and Engineering, Dhaka International University, Dhaka 1205, Bangladesh

Corresponding author: Md. Mahbubur Rahman (mahbub.rahman.cse@gmail.com)

ABSTRACT A brain tumor is a life-threatening neurological condition caused by the unregulated development of cells inside the brain or skull. The death rate of people with this condition is steadily increasing. Early diagnosis of malignant tumors is critical for providing treatment to patients, and early discovery improves the patient's chances of survival. The patient's survival rate is usually very less if they are not adequately treated. If a brain tumor cannot be identified in an early stage, it can surely lead to death. Therefore, early diagnosis of brain tumors necessitates the use of an automated tool. The segmentation, diagnosis, and isolation of contaminated tumor areas from magnetic resonance (MR) images is a prime concern. However, it is a tedious and time-consuming process that radiologists or clinical specialists must undertake, and their performance is solely dependent on their expertise. To address these limitations, the use of computer-assisted techniques becomes critical. In this paper, different traditional and hybrid ML models were built and analyzed in detail to classify the brain tumor images without any human intervention. Along with these, 16 different transfer learning models were also analyzed to identify the best transfer learning model to classify brain tumors based on neural networks. Finally, using different state-of-the-art technologies, a stacked classifier was proposed which outperforms all the other developed models. The proposed VGG-SCNet's (VGG Stacked Classifier Network) precision, recall, and f1 scores were found to be 99.2%, 99.1%, and 99.2% respectively.

INDEX TERMS Brain tumor, deep learning, detection, machine learning, MRI, prediction.

I. INTRODUCTION

A brain tumor is a clump of irregular cells in the brain that forms a mass [1]. The human brain is enclosed by a rigid skull. Any expansion in such a small area will trigger severe issues. Brain tumors can be cancerous and non-cancerous. The pressure within the skull will rise as benign or malignant tumors develop. This will result in permanent brain injury and even death. The Magnetic Resonance Imaging (MRI) picture of a healthy brain is shown in Figure 1 (A), while the picture of a brain containing a tumor is shown in Figure 1 (B). Approximately 700,000 people worldwide have a brain tumor, with approximately 86,000 new cases diagnosed in 2019. Since 2019, 16,830 people have died from brain tumors, with a 35 percent life expectancy [2]. As such, scientists and researchers have been working towards developing sophisticated techniques and methods for identifying brain tumors. Although MRI and Computer Tomography (CT) are the two modalities widely used for marking the abnormalities in terms of shape, size, or location of brain tissues which in turn help in detecting the tumors, MRI is preferred more by the doctors. As a consequence, scientists and researchers have more focused on MRI. While identifying brain tumors from MRI images, conventional inspection by physicians is mostly used. However, automated approaches mainly implemented by computer-aided medical image processing techniques are increasingly aiding physicians in detecting brain tumors.

Machine Learning (ML) algorithms gain insight from training data samples and can predict the class label of the unknown data objects. ML algorithms are popularly being used in the field of health-informatics [3]–[5], forecasting pandemic [6], evaluating user experience in playing games [7], predicting shear strength [8]. Similarly, many ML-based studies are conducted on medical images to classify brain tumors [9]–[11]. Medical image processing involves

The associate editor coordinating the review of this manuscript and approving it for publication was Haruna Chiroma.
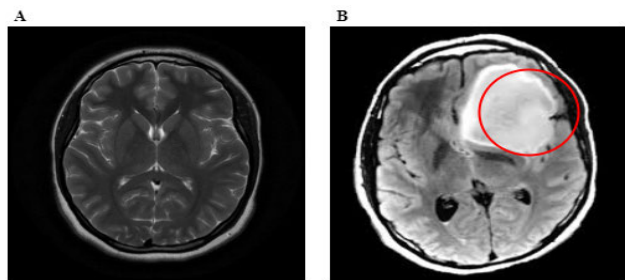
FIGURE 1. MRI Images: (A) healthy brain, (B) brain containing tumor.

pre-processing (enhancement, filter application, segmentation, feature selection) and post-processing (identification and/ classification) [12]. These steps can be implemented by the conventional machine learning approach as well as the deep learning approach. In the conventional machine learning approach, hand-crafted features are used to obtain results from test images and the process is fast. In the deep learning approach, models are tuned by appropriately selecting the number of layers, activation function, pooling, and sometimes pre-trained models are added for transfer learning. However, in both approaches' metaheuristic algorithms may be used to enhance the classification accuracy. From a broader perspective, this research covers both conventional and deep learning approaches for the identification of brain tumors from MRI images. A significant body of research has been conducted focusing on the detection of brain tumors using MRI Images through machine learning approaches. Although the conventional machine learning approach is faster in comparison to deep learning, the accuracy of the deep learning approach is better than the conventional machine learning approach.

The objectives of this research are, firstly, to explore how various image processing techniques are applied on MRI images for the detection of brain tumors. Secondly, to compare the performance of existing image processing techniques applied on MRI images for the detection of brain tumors. Finally, to propose an efficient technique for the detection of brain tumors using MRI images through machine learning approaches. As a result, this research will provide the expected outcome i.e., efficient image processing technique to detect brain tumors using MRI images through machine learning approach which will assist pathology experts to provide proper treatment.

Later sections of this paper are organized as follows. The related works are briefly introduced in Section 2. Section 3 briefly discusses developing and analyzing ML models. Later on, in Section 4, the performance of the ML models is compared. Finally, the discussion and conclusion are stated in Section 5.

II. LITERATURE REVIEW
This section briefly discusses the studies that are conducted to detect brain tumors using different state-of-the-art technologies. Rehman *et al.* [2] proposed a new learning-based

method for microscopic brain tumor detection and tumor type classification. The first phase of their study was to build a 3D convolutional neural network (CNN) architecture to extract brain tumors, which are then transferred to a CNN model that has already been trained to extract features. The extracted features are fed into a correlation-based selection process, and the best features are chosen as the result. For final classification, these selected features are tested using a feed-forward neural network. Amin *et al.* [13] employed a Deep Neural Network (DNN) based architecture for brain tumor segmentation. For classification, the proposed model employs 07 layers, including 03 convolutional, 03 ReLU, and a SoftMax layer. Thaha *et al.* [14] proposed a deep learning method using CNN for the segmentation. This method employs 3×3 small kernels for the deep architecture of the CNN model. Intensity normalization and data augmentation have been performed for the preprocessing of images. Kebir *et al.* [15] proposed a supervised method for detecting the brain abnormalities from the MRI images in three steps, first step is to develop a deep learning CNN model, then a subdivision of brain MRI images is done by the k-mean algorithm followed by brain component classification as normal or abnormal classes according to the developed CNN model. Vinoth *et al.* [16] proposed a programmed division strategy based on CNN. Here, kernels are used for classification, and SVM classification is performed with the calculated parameters. And, extraction and detection of tumors from MRI scan images of the brain are done by using the MATLAB tool. A three Incremental Deep Convolutional Neural networks 2CNet, 3CNet, and EnsembleNet for automatic brain tumor segmentation have been proposed in [17]. This method adopted the technique of Ensemble Learning and to avoid the hit and trial for training the CNN, they bounded the hyper-parameters to accelerate the training. Mohsen *et al.* [18] used DNN for classifying a dataset of 66 brain MRIs into 4 classes (normal, glioblastoma, sarcoma, and metastatic bronchogenic carcinoma tumors). A classifier was combined with the DWT and PCA. An automatic brain tumor segmentation algorithm [19] has been proposed using a Deep Convolutional Neural Network. An effective brain tumor segmentation from MRI images has been proposed in [20] by extracting the relevant features from combining the segmented pathological tissues, white matter, gray matter, and fluid (CSF) and then classifying them using the Neural Network model. The comparison has been done by implementing the k-nearest neighbor classifier and Bayesian Classifier.

In summary, it can be said from the literature review that: (a) None of the existing approaches are fully automated as calibration of processing parameters is essential. (b) If one batch of images works with one specific parameter new batch of images requires modification of parameters again. (c) None of the existing approaches have validated their proposed approach with other existing approaches. (d) The accuracy rate does not meet the ''gold standard'' criteria. (e) Human intervention is essential. Thus, more research in
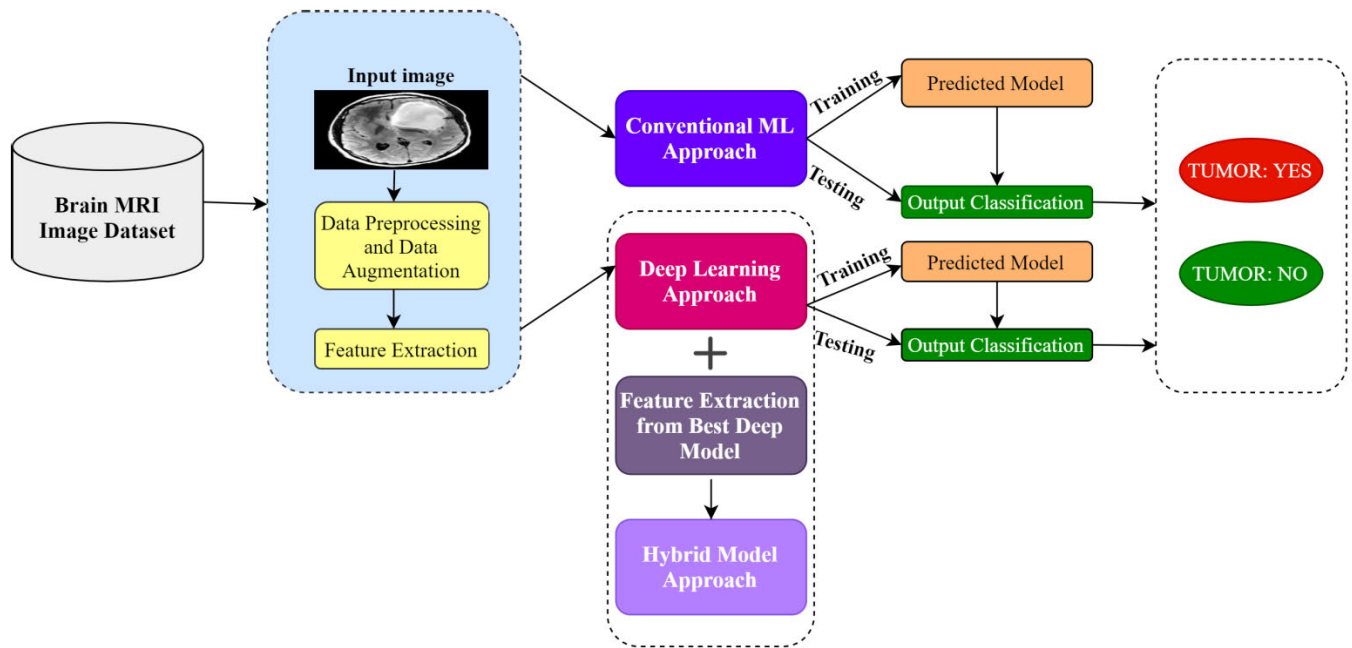
FIGURE 2. The workflow diagram of the brain tumor detection study.

this field needs to be carried out in order to fill the research gaps.

III. METHODOLOGY

The research was conducted in three phases: First, different traditional machine learning (ML) models are developed to find the best algorithm for detecting brain tumors. The traditional algorithms were selected based on the literature review of the previous works and the selected algorithms include Convolutional Neural Network (CNN), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), Naive Bayes (NB) and K Nearest Neighbors (KNN). In the second phase, fourteen different transfer learning models are developed. To do that, CNN model from the first phase is used as the base CNN model and fourteen different pre-trained models are used on the top layer of the CNN model keeping the target to find out the best pre-trained model. In this phase, for developing each of the transfer learning models, initial weights from the pre-trained models are obtained and these initial weights are used to initialize the base CNN model so that the model can be trained efficiently with a minimum number of epochs. In the last phase, four different hybrid models are proposed to detect brain tumors. For developing each of the hybrid models, features were extracted from the top-performing transfer learning model. Then, the extracted features were used as input/ independent variables for building the hybrid models. The algorithms used in this phase include Stacked Classifier (SC), AdaBoost (AB), CatBoost (CB), and XgBoost (XB). One of the main objectives of analyzing different machine learning algorithms was to find out the top-performing algorithm for brain tumor classification. To achieve this, we thoroughly carried out the

following steps: data collection, data synthesis, development of prediction models. These steps are discussed in detail, as follows. The workflow diagram of this research is shown in Figure 2.

A. DATA COLLECTION

In this research, an open-access dataset, which is available on Kaggle, was used for training the models. The open-access dataset contains a total of 253 MRI Images where 140 images were labeled as YES whether the rest of the images were classified as NO. However, for testing the ML models, a separate dataset containing 90 images of samples was used, which was collected from a renowned Pathology Institute in Bangladesh. Additionally, one of the medical experts from the same pathology laboratory, who has 30 years of experience in histology and serving as an Associate Professor in a Medical College, was nominated as the domain expert for this research study. The test dataset was meticulously labeled by the nominated domain expert.

B. DATA SYNTHESIS

In the second step, the collected images from the first step were synthesized. First, duplicate images were removed from the dataset. Second, completely black portions of each of the images were removed. To do that for each of the images, contours are identified on the top, bottom, left, and right direction based on the presence of the black regions. Each image was cropped based on these four contours. Hence, any portion out of these contours is removed from the images. Thus, the cropped image set contains the region of interest for each of the images. There was a class imbalance problem, the number of yes and no in the dataset was not the same,
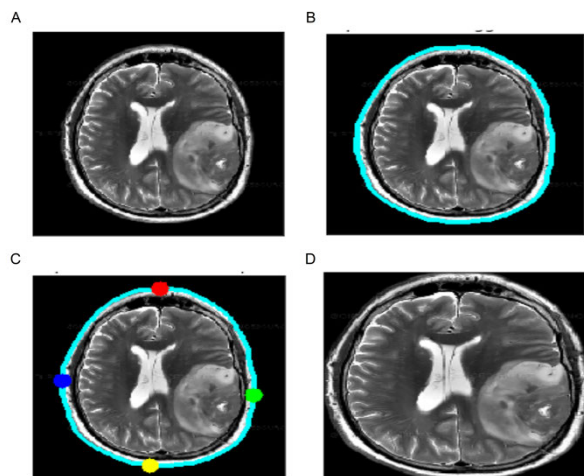
FIGURE 3. Steps in cropping a particular image: (A) original image, (B) finding contours, (C) finding extreme points, (D) cropped image.

in the combined dataset. Therefore lastly, data augmentation was done to handle this issue. As the number of images labeled as 'no' was less than the number of images labeled as 'yes', for each image, which was labeled as 'yes', 10 new images were generated, on the other hand for each image which was labeled as 'no', 7 new images were generated. Figure 3 shows the steps required in cropping a specific image data.

C. CONVENTIONAL ML MODEL DEVELOPMENT

In this step, different algorithms are chosen based on the recent works relating to brain tumor classification, which included CNN, SVM, RF, DT, NB, and KNN. The models were developed using scikit-learn, which is a Python module integrated into a wide range of machine learning algorithms. Each of the algorithms and their working procedure is analyzed in detail in the following sub-sections.

1) CONVOLUTIONAL NEURAL NETWORK (CNN)

Due to the secret potential to use the geometric of the images, CNN's main applications are in the field of image processing [20]. CNN outperforms many strategies in graph analysis [21]. It combines three architectural ideas: local receptive fields, shared weights, and spatial or temporal subsampling. The architecture of the proposed CNN model is shown in Figure 4. The proposed CNN structure for the study holds a total of seven layers which are as follows. First, the input layer: image features are given as input in this layer. Second, convolutional layer: 32 filters/ layer was used while having a 7×7 size kernel. Basically, in a convolutional layer, a filter is applied to the images and unnecessary details are removed while keeping the relevant information. Nonetheless, the sigmoid was used as the activation function in this layer. Third, max pooling layer: max pooling is performed in this layer with a pool size of 4×4. Fourth: dropout layer with a dropout rate of 50%, this layer removed 50% neurons randomly from
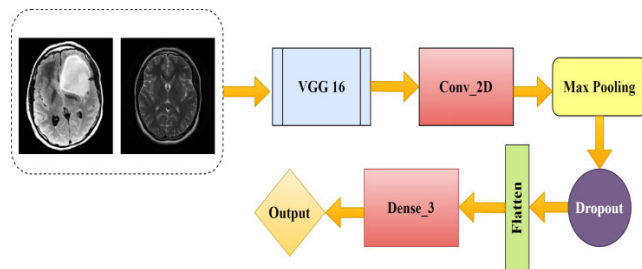
FIGURE 4. The architecture of the proposed CNN model.

the whole neural network architecture. Fifth, flatten layer: which is the function that converts the pooled feature map to a single column that is passed to the fully connected layer. Sixth, dense layer, which is a fully connected layer where the features from the previous layer are given as input. A sigmoid function was used as an activation function for this layer. Seventh, output layer, which gives the prediction probability that whether a particular MRI image contains a tumor or not. For the prediction probability is more than 50%, a particular image was considered to be containing a brain tumor else the image was considered not containing a brain tumor.

2) SUPPORT VECTOR MACHINE (SVM)

SVM is a supervised machine learning algorithm that can be used to solve classification and regression problems [22]. This algorithm creates a hyperplane (or a group of hyperplanes) in a high- or infinite-dimensional space to determine the best boundary between the potential outputs. In general, the aim is to find a hyperplane in *n-dimensional* space that maximizes the isolation of data points from their possible groups. SVMs can accommodate these wide feature spaces because they use overfitting protection that isn't dependent on the number of features.

3) RANDOM FOREST (RF)

The RF classifier is an ensemble approach that uses bootstrapping and aggregation to train multiple decision trees in parallel, the process known as bagging [23].

4) DECISION TREE (DT)

DT is a well-known hierarchical machine learning method for the prediction that employs a tree-like model of decisions and their potential outcomes [24]. Each internal node (not a leaf node) of DT evaluates an attribute, each branch represents the test results, and each leaf node (or terminal node) specifies the class label.

5) NAÏVE BAYES (NB)

Naive Bayes is a classification algorithm that is both supervised and statistical in nature [25]. It predicts the class of an undefined data set using Bayes' probability theorem. It calculates membership probabilities for each class, such as the probability that a given record or data point belongs to a

certain one. The most possible class is the one with the greatest probability.

6) K NEAREST NEIGHBOR (KNN)

KNN is a supervised machine learning algorithm that predicts the values of new data points using "feature similarity" [26]. This stores the feature vectors and class labels of the training samples in the training phase, and the unlabeled sample is categorized in the classification phase by assigning the class label based on the most frequent of the k training samples closest to the query point, where k is a defined constant.

D. DEEP ML MODEL DEVELOPMENT

In this section, fourteen different pre-trained models are used to build fourteen different transfer learning models. Transfer learning is a machine learning technique in which a model created for one job is utilized as the basis for a model on a different task. The pre-trained models are used to initialize the weights of the CNN model in this research include: VGG-16, VGG-19, Xception, ResNet152, ResNet50V2, ResNet101V2, InceptionV3, InceptionResNetV2, MobileNet, MobileNetV2, DenseNet121, DenseNet169, DenseNet201 and NASNetMobile. Among the pre-trained models, VGG-16 was found to be the best performing, which is analyzed in detail in section IV. The overview of the VGG-16 pre-trained model is provided in the following subsection.

1) VGG16

VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition". The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was one of the famous models submitted to ILSVRC-2014. It improves AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3 × 3 kernel-sized filters one after another. The architecture of the VGG-16 model is presented in Figure 5.

E. HYBRID MODEL ANALYSIS

In this section, different hybrid models are developed for detecting brain tumors. To develop the hybrid models, the features were extracted from the second last (sixth) layer of the top-performing transfer learning model (VGG16_CNN) as it provides the second most reduced set of features. In total there were 32 features in that layer. These features were then used as input for developing four other models, such as AdaBoost, CatBoost, XGboost, and the Stacked classifier. However, due to high time and space complexity, developing these models considering all the features of the images was not feasible. Thus, the VGG16_CNN model was used as a feature selection method to gain a reduced feature set, which is representative of the whole feature set. Nonetheless, these models are analyzed in detail in the following subsections.
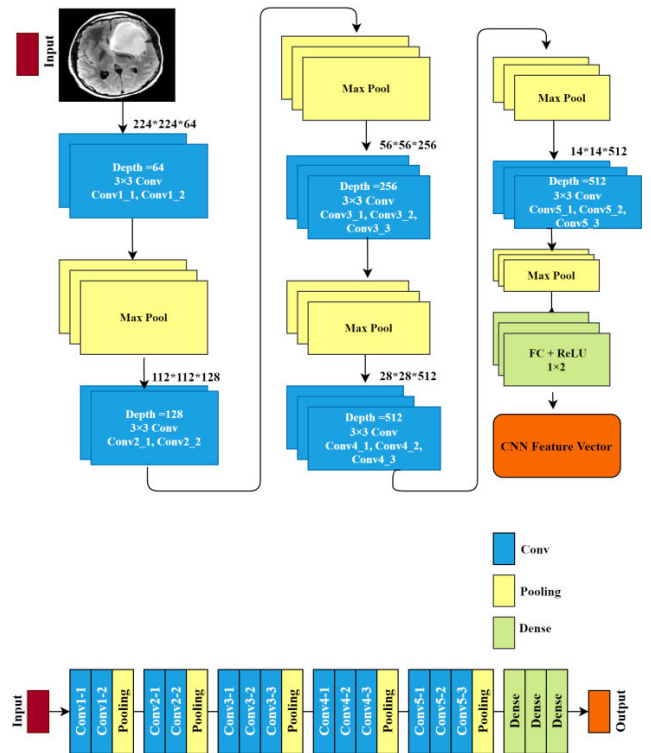

FIGURE 5. Architecture of the VGG16 model.

1) ADABOOST

AdaBoost, short for Adaptive Boosting, is an AI meta-calculation that can be utilized related to numerous different sorts of learning calculations to improve execution. The yield of the other learning calculations ('feeble learner') is consolidated into a weighted whole that addresses the last yield of the supported classifier. AdaBoost is versatile as in ensuing feeble learner are changed for those occasions misclassified by past classifiers. In certain issues, it very well may be less vulnerable to the overfitting issue than other learning calculations. The individual learner can be feeble, however as long as the presentation of everyone is somewhat better compared to arbitrary speculating, the last model can be demonstrated to join to a solid learner.

2) CATBOOST

CatBoost is an open-source programming library created by Yandex. It gives a slope-boosting structure that endeavors to tackle categorical highlights utilizing a stage-driven option contrasted with the old-style algorithm. CatBoost has acquired ubiquity contrasted with other angle boosting calculations fundamentally because of the accompanying highlights: (a) Ordered Boosting to defeat overfitting. (b) Native dealing with clear-cut highlights. (c) Using Oblivious Trees or symmetric trees for quicker execution.

3) XGBOOST

XGBoost is a decision tree-based ensemble Machine Learning algorithm built on a gradient boosting framework. Artificial neural networks outperform all other algorithms or
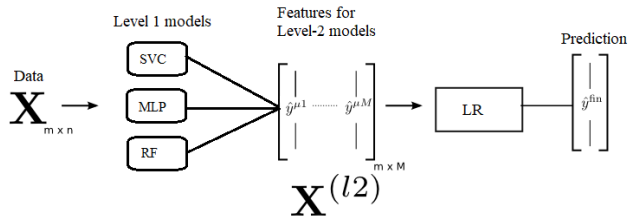
FIGURE 6. The architecture of the proposed stacked classifier.

TABLE 1. System specifications.

System Specification for the Model Development	
RAM	13GB
CPU	1×single core hyper threaded i.e. (1 core, 2 threads) Xeon Processors @2.3Ghz
Cache	46MB
GPU	NVidia K80 GPU
GPU Memory	16GB
Session Limit	9 hours
Disk Space	100GB

systems in prediction problems involving unstructured data (images, text, etc.). However, decision tree-based algorithms are currently considered best-in-class for small-to-medium structured/tabular data.

4) STACKED CLASSIFIER

A 2-layer stacked classifier was built on 32 features, which was selected by the neural network model. The architecture of the proposed stacked classifier is shown in Figure 6.

Three basic ML algorithms, namely SVM, Multi-Layer Perceptron (MLP), and RF were kept as the first layer of the SC model whereas, a Logistic Regression (LR) model was used in the second layer of the stacked classifier model. For each of the observations/ samples in the dataset, the verdicts are obtained from three different algorithms. The verdicts obtained from these algorithms are used as input for the second layer LR model. Then, the final verdict was obtained from the second layer model.

IV. RESULTS

To evaluate the performance of each prediction model, a separate test set was used which contains a total of 90 images, where 81 images were labeled as YES, on the other hand, the rest of the images were labeled as NO. The performance of each prediction model was measured in terms of its precision, recall, and F1 scores. The experiments have been conducted on a local machine platform, which provides the following specifications (Table 1).

Each of the evaluation parameters was obtained by performing macro averaging on the actual and model-predicted class labels, which calculated parameters for each class label
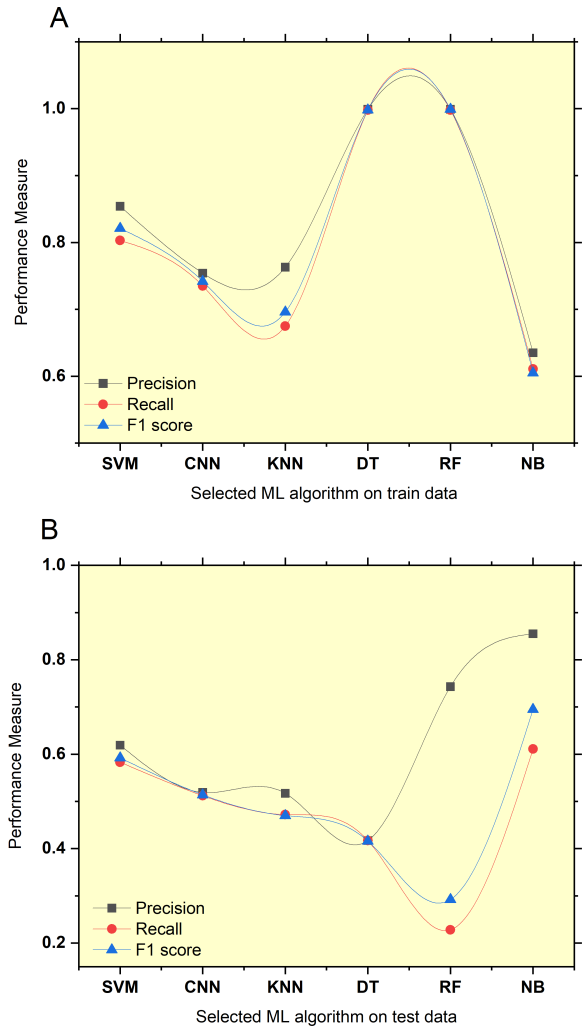

FIGURE 7. 7(A) represents the performance of the ML-based algorithms on the train dataset whereas, 7(B) represents the performance on the test dataset.

and found their unweighted means. Since the resampling method had been utilized to balance the classes, accuracy was not considered a performance metric for evaluating the performance of the classifiers as the literature had shown that accuracy was not an appropriate metric to use in such a case [27]. The performance of running each of the conventional machine learning algorithms is presented in Figure 7.

The best training performance was obtained by CNN, RF, DT, and KNN followed by SVM and DT (see Figure 7). However, the best test performance was obtained by CNN, followed by SVM, RF, DT, NB, and KNN. From the selected algorithms, CNN achieved 100% train performances for different performance metrics including precision, recall, and f1-score, while it obtained an 88.7% precision, recall, and f1-score for each in analyzing the performance on the test dataset. For the SVM algorithm, the precision, recall, and f1-score for the training dataset were 97.9%, 96%, and 96% respectively whereas, the precision, recall, and f1 score for

TABLE 2. Results obtained from the deep CNN architecture.

Deep CNN Model	Train			Test		
	Precision	Recall	F1 score	Precision	Recall	F1 score
VGG-16	1	1	1	0.978	0.978	0.978
VGG-19	1	1	1	0.941	0.944	0.942
Xception	0.811	0.717	0.684	0.847	0.878	0.862
ResNet152	0.521	0.464	0.325	0.785	0.178	0.221
ResNet50V2	0.293	0.541	0.38	0.85	0.922	0.885
ResNet101V2	0.865	0.824	0.822	0.938	0.7	0.77
InceptionV3	0.833	0.809	0.808	0.926	0.767	0.819
InceptionResNetV2	0.854	0.813	0.811	0.901	0.656	0.736
MobileNet	0.781	0.582	0.515	0.874	0.244	0.307
MobileNetV2	0.754	0.469	0.31	0.928	0.1	0.055
DenseNet121	0.813	0.708	0.691	0.933	0.5	0.598
DenseNet169	0.867	0.84	0.839	0.937	0.656	0.735
DenseNet201	0.828	0.726	0.711	0.896	0.6	0.691
NASNetMobile	0.756	0.574	0.508	0.874	0.244	0.307

the test dataset were 94.6%, 85.4%, and 89.7%. The RF algorithm also had a promising performance on the training dataset as the value of all the evaluation parameter were 100%. The test performance of RF was 86.6%, 80%, and 82.8% for the precision, recall, and f1 scores respectively. Also, For the DT model, values of all the evaluation parameters for the training dataset were 100% whereas, the values of the evaluation parameters for test data for precision, recall, and f1 score were 88.5%, 73.3%, and 78.7% respectively. The train performance was lowest for NB algorithm among all the algorithms as the values of precision, recall and f1-score were 63.5%, 61.1%, and 60.5%. However, the value of the precision, recall, and f1 scores for the test dataset considering this algorithm were 85.5%, 61.1%, and 69.5%. Again, the result for the KNN algorithm shows that precision, recall, and f1-scores for the training dataset were 100% whereas the precision, recall, and f1 score for the test dataset were 84.7%, 55.6%, and 65%. It can be observed that the lowest f1 score has been obtained by the KNN algorithm.

The evaluation result for the deep ML-based pre-trained models is presented in Figure 10, while all the scores obtained from those models are shown in Table 2. It can be observed from Figure 10 that the best performance for both the train as well as the test dataset was obtained by the VGG-16 pre-trained model. The precision, recall, and f1 score of the VGG-16 model on the train dataset was 100%, while the precision, recall and f1 score on the test dataset was 97.8%. The confusion matrix for the test and train data of the VGG-16 model is highlighted in Figure 8, while the accuracy vs. loss curve for each epoch of the model is shown in Figure 9.

Again, for the final phase of this research, the performance of the hybrid models on the test data is represented in Table 3. Based on the results it is evident that Stacked Classifier (SC) achieves the best score and outperforms others, available hybrid model. However, precision, recall, and f1 score for the AdaBoost model were 94.2%, 93.3%, and 93.7%. The performance measures for the CatBoost model were 93.9%,
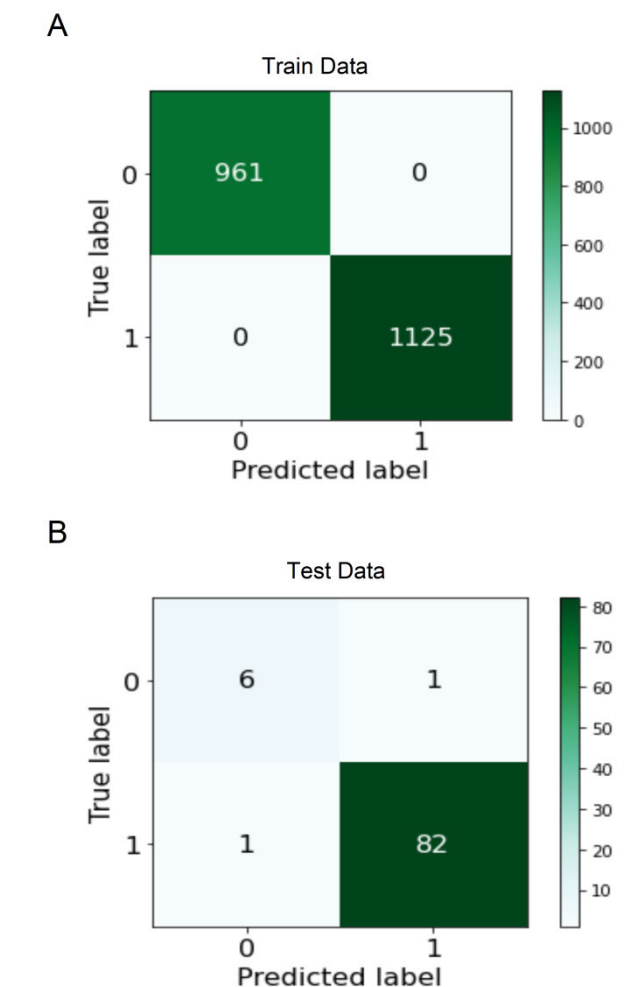
FIGURE 8. For VGG16 architecture 9(A) shows the confusion matrix for train data whereas, 9(B) represents the confusion matrix for test data.

94.4%, and 93.9%. And, for the XgBoost model, precision, recall, and F1 scores were 95.6%. Thus, the best performance was obtained by the SC model with 99.1%, 98.9%, and 99.2% of precision, recall, and f1 scores respectively.
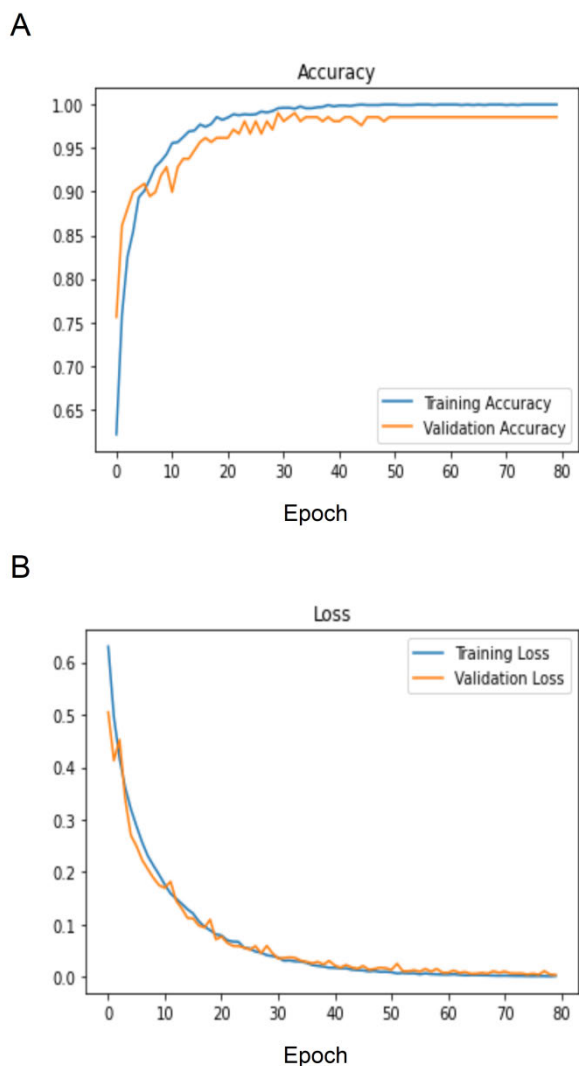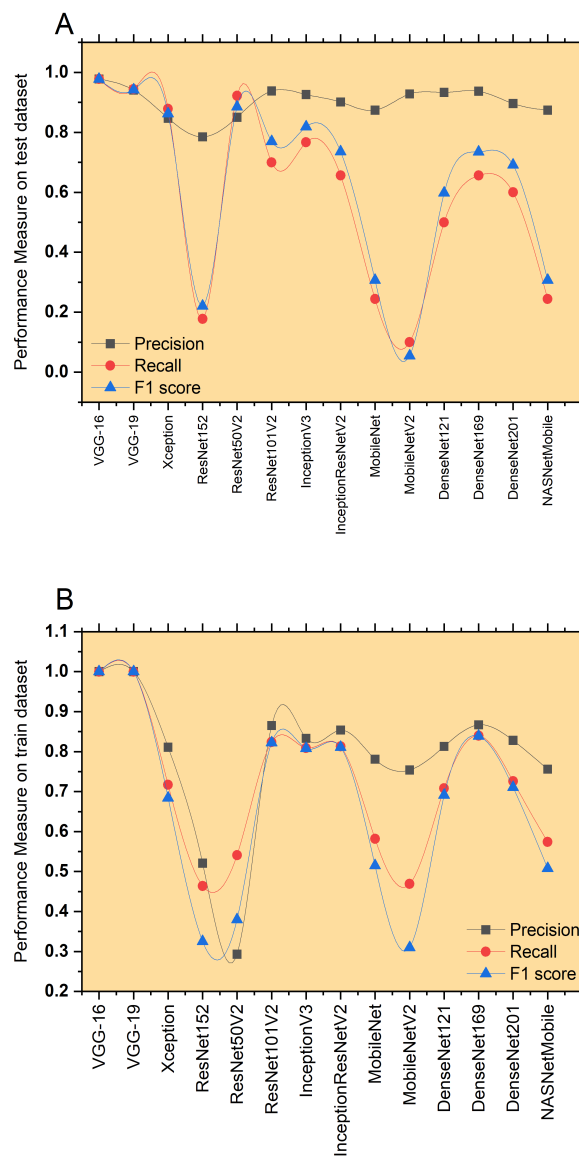
FIGURE 9. For each epoch of VGG16 architecture 10(A) shows the curve of Train vs Validation accuracy and 10(B) represents the Train vs Validation loss curve.

TABLE 3. Results obtained from the hybrid models.

Model	Precision	Recall	F1 score
Stacked	0.992	0.991	0.992
AdaBoost	0.942	0.933	0.937
CatBoost	0.939	0.944	0.939
XGBoost	0.956	0.956	0.956

FIGURE 10. The test performance for the deep CNN-based models is shown in 10(A) whereas, the train performance is shown in 10(B).

V. DISCUSSION

This research yielded three clearly defined outcomes. First, the best performing ML model was identified for classifying brain tumors from MRI images by applying different classical algorithms, subsuming: CNN, SVM, RF, DT, NB, and KNN, on the labeled dataset. It was found that CNN with seven fine-tuned layers gives better performance (88.7% F1 Score) than other ML algorithms. Second, among 14 different pre-trained DL models, the top-performing pre-trained model was identified by having these pre-trained models on top

of the proposed CNN architecture. In this study, VGG-16 (F1 Score 97.8%) was found as the best pre-trained model for classifying brain tumor images. Third, four different hybrid models are proposed by extracting features from the second last layer of the VGG-16_CNN transfer learning model. The proposed Stacked Classifier (SC) hybrid model provides the best performance than all the other models.

While carrying out the study, the following issues are identified: firstly, there is no Benchmark Dataset that can be used to compare existing approaches for the detection of brain tumors from MRI images. Thus, there is a scope of data collaboration. The dataset made available for this study may be used as Benchmark Dataset (with prior permission from the originator). Secondly, a huge volume of

TABLE 4. Overview of research using deep learning approaches with their working procedure and performance metrics for Brain tumor classification.

Reference	Modality	Method	Remarks	Performance Metrics and Results
Mohsen et al. [18]	Classification of brain tumors	DNN	A dataset of 66 brain MRIs was classified into four classifications using a DNN classifier, including normal, glioblastoma, sarcoma, and metastatic bronchogenic carcinoma tumors. The classifier was paired with the discrete wavelet transform (DWT), a powerful feature extraction method, and principal component analysis (PCA), and the overall performance was quite well.	Classification Rate: 96.97%, Recall 97%, Precision 97%, F1 Score 97%, AUC-98.4%
Hu et al. [28]	Classification of brain tumors	CNN	For the best feature selection and picture classification, an updated version of the seagull optimization technique is used. The proposed method's simulation results are compared to those of a few other methods. When compared to the other approaches, the final findings show that the suggested method outperforms them in terms of CDR, FAR, and FRR indices.	CDR (Correct Decision Rate): 86%, FAR (False Acceptance Rate) 8%, FRR (False Rejection Rate): 6%
Rao et al. [29]	Brain tumor segmentation	DNN	The results of using DNNs to segment brain tumors for the BRATS 2015 challenge are presented in this publication. A pixel-wise classification method is used to detect malignancies in brain imaging.	Accuracy (AC): 67%
Ari and Hanbay [30]	Brain tumor segmentation	ELM-LRF	The study's goal was to use only cranial MR images with a mass in order to save time for the doctors.	Accuracy (AC): 97.18 %
Nie et al. [31]	Imaging-guided survival time prediction of brain tumor patients	3D CNN	Presented deep learning frameworks for autonomously extracting characteristics from high-grade glioma patients' multi-modal preoperative brain imaging (i.e., T1 MRI, fMRI, and DTI).	Accuracy 89.85 %, sensitivity 96.87%, Specificity 83.9%
Saba et al. [32]	Brain tumor detection	VGG-19	The Grab cut approach is used to segment real lesion symptoms, while the Transfer learning model visual geometry group (VGG-19) is fine-tuned to acquire features that are then concatenated with hand-crafted (shape and texture) features using a serial-based technique.	Dice Similarity Coefficient 99%
Chanu et al. [33]	Brain tumor classification	2D CNN	Classified images into two classes (normal and infected)	Accuracy with 97%, sensitivity with 100%, and Specificity with 94%.
Thaha et at. [14]	Brain tumor segmentation	CNN	Proposed a deep learning method using CNN for the segmentation. a) The method employs 3×3 small kernels for the deep architecture of the CNN model. b) Intensity normalization and data augmentation have been performed for the preprocessing of image	complete, core, and enhancing regions in Dice Similarity Coefficient metric (0.88, 0.83, 0.77) f
Proposed model (VGG-SCNet)	Detection of brain tumor	ML+ Deep CNN +Stacked Classifier	Stacking Classifier (SC) hybrid model provides the best performance over other models	Precision:99.2%, recall:99.1%, and F1 score:99.2%

data was not available which is a prerequisite for developing the deep learning models. Thirdly, extensive hyperparameter tunning can help to obtain better performance for ML and DL models as by doing so better performances are obtained for the developed models in this study. Fourthly, there is a trade-off between the algorithmic performance and the time complexity; better performances are obtained by the DL techniques, where the time complexity of these techniques was very high, meaning that it required a colossal amount of time to obtain the results on these techniques. On the other hand, the time complexity of the ML algorithms was comparatively low. Fifthly, the performance of classification largely

depends on the dataset and techniques that are being adopted; different performances can be obtained on the same dataset while applying different state-of-the-art methodologies; the same methodology provides different performances for the different datasets. Nonetheless, a comprehensive overview of similar research outcomes with the proposed model has been further designed to understand the state-of-the-art methodologies and their performance metrics for the classification of Brain Tumor (Table 4).

This research has the following limitations. First, the number of images considered in this study contains a limited number of images where using more images would make the study more robust. Second, no traditional image processing-based classification techniques were analyzed in this study. Third, the current study only classifies brain tumors from 2D data. Fourth, this study only focuses on classifying MRI images into tumorous and non-tumorous, no grading of tumors (e. g. Grade 1, 2, 3, 4) is done which helps the clinicians to determine its size, whether it has spread and the best treatment options available. Therefore, future studies may include: (a) working on both ML and image processing techniques with more images for classifying brain tumors from MRI images, (b) classifying brain tumors can be done on both on 2D all as 3D data, (c) grading of tumors can be obtained if the colossal amount of data is available (collaboration with hospitals may ensure the availability of huge volume of data), and (d) analyzing the computational complexity of the proposed models. Furthermore, a Benchmark Dataset can be contributed by the appropriate Medical Authority for carrying out comparisons of the existing approaches (classical image processing, and ML-based classification methods).

VI. CONCLUSION

MRI-based medical image analysis for brain tumor studies has been gaining attention in recent times due to an increased need for efficient and objective evaluation of large amounts of medical data. Because of the high death rate linked with brain tumors, it is critical to diagnose them early to treat them and reduce mortality. Manual diagnosis of the brain and tumor tissues is time-consuming and operator-dependent due to the intricacy of brain tissue. Therefore, in this research, an effective transfer learning-based SC model, which was named *VGG-SCNet*, was proposed to classify brain tumors from MRI Images. The F1 score obtained by the proposed classifier shows the efficacy of the approach followed by this study.

REFERENCES

[1] *Brain Tumor*. Accessed: Apr. 11, 2021. [Online]. Available: https://www.healthline.com/health/brain-tumor

[2] A. Rehman, M. A. Khan, T. Saba, Z. Mehmood, U. Tariq, and N. Ayesha, "Microscopic brain tumor detection and classification using 3D CNN and feature selection architecture," *Microsc. Res. Techn.*, vol. 84, no. 1, pp. 133–149, 2021.

[3] M. N. Islam, T. Mahmud, N. I. Khan, S. N. Mustafina, and A. K. M. N. Islam, "Exploring machine learning algorithms to find the best features for predicting modes of childbirth," *IEEE Access*, vol. 9, pp. 1680–1692, 2021.

[4] N. I. Khan, T. Mahmud, M. N. Islam, and S. N. Mustafina, "Prediction of cesarean childbirth using ensemble machine learning methods," in *Proc. 22nd Int. Conf. Inf. Integr. Appl. Services*, Nov. 2020, pp. 331–339.

[5] A. I. Aishwarja, N. J. Eva, S. Mushtary, Z. Tasnim, N. I. Khan, and M. N. Islam, "Exploring the machine learning algorithms to find the best features for predicting the breast cancer and its recurrence," in *Proc. Int. Conf. Intell. Comput. Optim.* New York, NY, USA: Springer, 2020, pp. 546–558.

[6] M. N. Islam and A. N. Islam, "A systematic review of the digital interventions for fighting COVID-19: The Bangladesh perspective," *IEEE Access*, vol. 8, p. 114078–114087, 2020.

[7] T. Zaki, N. I. Khan, and M. N. Islam, "Evaluation of user's emotional experience through neurological and physiological measures in playing serious games," in *Proc. Int. Conf. Intell. Syst. Des. Appl. (ISDA)*. New York, NY, USA: Springer, 2021, pp. 1039–1050.

[8] J. Rahman, K. S. Ahmed, N. I. Khan, K. Islam, and S. Mangalathu, "Data-driven shear strength prediction of steel fiber reinforced concrete beams using machine learning approach," *Eng. Struct.*, vol. 233, Apr. 2021, Art. no. 111743.

[9] P. Kaur, G. Singh, and P. Kaur, "Classification and validation of MRI brain tumor using optimised machine learning approach," in *Proc. ICDSMLA*. New York, NY, USA: Springer, 2020, pp. 172–189.

[10] T. Pandiselvi and R. Maheswaran, "Efficient framework for identifying, locating, detecting and classifying MRI brain tumor in MRI images," *J. Med. Syst.*, vol. 43, no. 7, pp. 1–14, Jul. 2019.

[11] A. Tiwari, S. Srivastava, and M. Pant, "Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019," *Pattern Recognit. Lett.*, vol. 131, pp. 244–260, Mar. 2020.

[12] I. Bankman, *Handbook of Medical Image Processing and Analysis*. Amsterdam, The Netherlands: Elsevier, 2008.

[13] J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes, "Big data analysis for brain tumor detection: Deep convolutional neural networks," *Future Gener. Comput. Syst.*, vol. 87, pp. 290–297, Oct. 2018.

[14] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016.

[15] S. T. Kebir and S. Mekaoui, "An efficient methodology of brain abnormalities detection using CNN deep learning network," in *Proc. Int. Conf. Appl. Smart Syst. (ICASS)*, Nov. 2018, pp. 1–5.

[16] R. Vinoth and C. Venkatesh, "Segmentation and detection of tumor in MRI images using CNN and SVM classification," in *Proc. Conf. Emerg. Devices Smart Syst. (ICEDSS)*, Mar. 2018, pp. 21–25.

[17] R. Saouli, M. Akil, M. Bennaceur, and R. Kachouri, "Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in MRI images," *Comput. Methods Programs Biomed.*, vol. 166, pp. 39–49, Nov. 2018.

[18] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Comput. Informat. J.*, vol. 3, no. 1, pp. 68–71, 2018.

[19] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.

[20] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, "Convolutional neural network (CNN) for image detection and recognition," in *Proc. 1st Int. Conf. Secure Cyber Comput. Commun. (ICSCCC)*, Dec. 2018, pp. 278–282.

[21] M. Henaff, J. Bruna, and Y. LeCun, "Deep convolutional networks on graph-structured data," 2015, *arXiv:1506.05163*. [Online]. Available: http://arxiv.org/abs/1506.05163

[22] O. L. Mangasarian and D. R. Musicant, "Lagrangian support vector machines," *J. Mach. Learn. Res.*, vol. 1, pp. 161–177, Sep. 2001.

[23] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[24] R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 1993.

[25] K. P. Murphy, "Naive Bayes classifiers," *Univ. Brit. Columbia*, vol. 18, no. 60, pp. 1–8, Oct. 2006.

[26] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007.

[27] N. V. Chawla, "Data mining for imbalanced datasets: An overview," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Boston, MA, USA: Springer, 2010, doi: 10.1007/978-0-387-09823-4_45.

[28] A. Hu and N. Razmjooy, "Brain tumor diagnosis based on metaheuristics and deep learning," *Int. J. Imag. Syst. Technol.*, vol. 31, no. 2, pp. 657–659, Jun. 2021.

[29] V. Rao, M. S. Sarabi, and A. Jaiswal, "Brain tumor segmentation with deep learning," in *Proc. MICCAI Multimodal Brain Tumor Segmentation Challenge*, Oct. 2015, pp. 1–4.

[30] A. Ari and D. Hanbay, "Deep learning based brain tumor classification and detection system," *Turkish J. Elect. Eng. Comput. Sci.*, vol. 26, no. 5, pp. 2275–2286, Sep. 2018.

[31] D. Nie, H. Zhang, E. Adeli, L. Liu, and D. Shen, "3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Oct. 2016, pp. 212–220.

[32] T. Saba, A. S. Mohamed, M. El-Affendi, J. Amin, and M. Sharif, "Brain tumor detection using fusion of hand crafted and deep learning features," *Cogn. Syst. Res.*, vol. 59, pp. 221–230, Jan. 2020.

[33] M. M. Chanu and K. Thongam, "Computer-aided detection of brain tumor from magnetic resonance images using deep learning network," *J. Ambient Intell. Hum. Comput.*, vol. 20, pp. 1–2, Jul. 2020.

T. M. SHAHRIAR SAZZAD received the M.Sc. degree in information technology (IT) from the U.K., in 2007, and the Ph.D. degree in computer science engineering (CSE) from Australia, in 2017. His research interests include computer vision, pattern recognition, image processing, especially medical image processing, and human–computer interaction.

MOHAMMAD SHAHJAHAN MAJIB received the B.Sc. degree in computer science and engineering from Dhaka University (DU), Dhaka, Bangladesh, the M.Sc. degree in computer science and engineering from the Military Institute of Science and Technology (MIST), Dhaka, and the Postgraduate Diploma degree in wired communication engineering from the PLA University of Science and Technology (PLAUST), China. He is currently working as a Senior Instructor with the Department of Computer Science and Engineering, MIST. His research interests include medical image processing, pattern recognition, and machine learning.

NAFIZ IMTIAZ KHAN is currently a Lecturer with the Department of Computer Science and Engineering, Military Institute of Science and Technology (MIST), Dhaka, Bangladesh. He is the author of more than ten peer-reviewed publications in international journals and conferences. His research interests include machine learning, artificial intelligence, data science, image processing, edge computing, bioinformatics, and human–computer interaction (HCI).

MD. MAHBUBUR RAHMAN (Member, IEEE) received the Ph.D. degree in information computer science from Japan Advanced Institute of Science and Technology (JAIST), Japan, in 2004. He is currently a Running Professor with the Department of Computer Science and Engineering (CSE), Military Institute of Science and Technology (MIST), Bangladesh. He has authored more than 75 research articles in reputed journals and conferences globally. His research interests include computer networks, image processing, pattern recognition, bioinformatics, and machine learning.

SAMRAT KUMAR DEY received the B.Sc. degree in computer science and engineering from Patuakhali Science and Technology University (PSTU), Bangladesh, in 2015, and the M.Sc. degree in computer science and engineering from the Military Institute of Science and Technology (MIST), BUP, Bangladesh, in 2019. He is currently an Assistant Professor with the Department of Computer Science and Engineering (CSE), Dhaka International University (DIU), Bangladesh. He has published more than 30 research articles in reputed journals and conferences. His research interests include visual data analytics, health informatics, information visualization, HCI, machine learning, and deep learning. He has also served as a technical committee member and a reviewer for different international journals and conferences worldwide.

• • •