Vision Transformers, Ensemble Model, and Transfer Learning Leveraging Explainable AI for Brain Tumor Detection and Classification

Shahriar Hossain, Amitabha Chakrabarty, Thippa Reddy Gadekallu, *Senior Member, IEEE*, Mamoun Alazab, *Senior Member, IEEE*, and Md. Jalil Piran, *Senior Member, IEEE*

***Abstract*— The abnormal growth of malignant or non-malignant tissues in the brain causes long-term damage to the brain. Magnetic resonance imaging (MRI) is one of the most common methods of detecting brain tumors. To determine whether a patient has a brain tumor, MRI filters are physically examined by experts after they are received. It is possible for MRI images examined by different specialists to produce inconsistent results since professionals formulate evaluations differently. Furthermore, merely identifying a tumor is not enough. To begin treatment as soon as possible, it is equally important to determine the type of tumor the patient has. In this paper, we consider the multiclass classification of brain tumors since significant work has been done on binary classification. In order to detect tumors faster, more unbiased, and reliably, we investigated the performance of several deep learning (DL) architectures including Visual Geometry Group 16 (VGG16), InceptionV3, VGG19, ResNet50, InceptionResNetV2, and Xception. Following this, we propose a transfer learning(TL) based multiclass classification model called IVX16 based on the three best-performing TL models. We use a dataset consisting of a total of 3264 images. Through extensive experiments, we achieve peak accuracy of 95.11%, 93.88%, 94.19%, 93.88%, 93.58%, 94.5%, and 96.94% for VGG16, InceptionV3, VGG19, ResNet50, InceptionResNetV2, Xception, and IVX16, respectively. Furthermore, we use Explainable AI to evaluate the performance and validity of each DL model and implement recently introduced Vison Transformer (ViT) models and compare their obtained output with the TL and ensemble model.**

***Index Terms*— Brain Tumor Classification, Deep Learning, Ensemble Learning, Multiclass Classification, Transfer Learning, VGG16, VGG19, InceptionV3, Xception, ResNet50, InceptionResNetV2, Explainable AI, LIME, Vision Transformers, SWIN, CCT, EANet.**

I. Introduction

A brain tumor is characterized by the proliferation of brain cells. This disease is mainly caused by morbidity and cancer-related conditions [1]. Brain tumors are caused by malignant cells infiltrating brain tissues and growing abnormally [2]. In secondary brain tumors, tumors that originate elsewhere in the body and then spread to the brain are classified as secondary brain tumors [3]. The term metastatic brain cancer also applies to these cancers. In addition to lung, breast, skin (melanoma), colon, kidney, and thyroid gland cancers, the brain is a common site of metastasis for other types of cancer as well.

In the United States, there are approximately 787,000 people with brain tumor disorders, according to the National Brain Tumor Association [5]. MRI is a common imaging method used before and after surgery, as it provides important information for treatment plans [6], [4]. Consequently, it has led to significant advancements in the comprehensive study and in-depth phenotyping of the human brain [7]. With MRI imaging, soft tissues can be distinguished more clearly in three dimensions than with conventional imaging methods [8].

MRI scans have recently been analyzed using several machine learning (ML) techniques. Medical applications have demonstrated the potential of hybrid models, along with traditional DL models [49]. As an example, DL models, as a category of ML models, were extensively used to diagnose Covid-19 [9], [50]. A DL approach is employed to tackle pneumonia, utilizing both convolutional neural networks (CNN) [10] and recurrent neural networks (RNNs) [51]. The authors in [11] used some supervised and unsupervised methods to characterize tumors that originated in the lungs and pancreas, whereas [12] used DL to identify chest pathology. ML has also been applied to cognitive computing [13]. The ability of DL to process images effectively has been demonstrated for quite some time. The DL model, Alexnet [16], proved its capability to handle the enormous scale of image processing tasks. From that point forward, we have seen the passageway of a few incredible neural networks models like VGG16 [17] and remarkable achievements for image recognition in practical aspects like clinical analysis [18] in terms of lung disease detection; face recognition [19], DL models based on meta-learning [20], unmanned aerial vehicle imagery [21], etc.

Although MRI scans can be used manually to identify brain tumors, this is a static procedure that produces inaccurate results. DL algorithms can be used in these scenarios since they can replace manual methods that are time-consuming,

Shahriar and Amitabha are with the Department of Computer Science and Engineering, Brac University, Dhaka, Bangladesh, email: shahriar.hossain@bracu.ac.bd, amitabha@bracu.ac.bd

Gadekallu is with the School of Information Technology and Engineering, Vellore Institute of Technology, Vellore 632014, India as well as with the Department of Electrical and Computer Engineering, Lebanese American University, Byblos, Lebanon, email: thippareddy@ieee.org

Mamoud is with Charles Darwin University, Australia, email: alazab.m@ieer.org

Piran is with the Department of Computer Science and Engineering, Sejong University, South Korea, email: piran@sejong.ac.kr

laborious, and ineffective. For instance, in research [14] automatic segmentation and prediction of MRI scans are conducted. While in [23] Geometric Median Shift was used for the automatic detection and in [15] CNNs were used for automatic segmenting of brain tumors.

The aim of this study is to introduce an ensemble model based on the averaging of three other Transfer Learning(TL)-based models. Using a single model can sometimes lead to overfitting and biased conclusions. By using grouped weights, the average TL model overcomes these challenges. Six different TL models were implemented and three were selected for the ensemble model. As a result of averaging three models, the model becomes more complex due to a greater number of layers, making it capable of solving more complex problems and allowing better feature extraction. According to these results, the ensemble model produces more accurate classification and anomaly identification results than a single model. As a result of this research, professionals and nonprofessionals will be able to detect brain tumors from MRI images with greater accuracy.

Our main contributions are summarized as follows.

- Six CNN-based TL algorithms, including InceptionV3, VGG16, VGG19, ResNet50, InceptionResNetV2, and Xception, are implemented and their layers are tweaked to increase their effectiveness in terms of the classification of tumors.
- We propose the IVX16 ensemble model for detecting and classifying brain tumors in a multiclass mode.
- According to our extensive experiment results, IVX16 outperforms all three CNN-based TL algorithms in accuracy curves, confusion matrices, classification reports, and model explainability.
- To find out how our model compares to the SOTA models, we generate Local Interpretable Model-Agnostic Explanations (LIME) to check their validity.
- Furthermore, we implement three Vision-based transformer models named SWIN, CCT, and EANet. We then further compare their outputs with the aforementioned TL models.

The remainder of the paper is organized as follows. In Section II, related work is discussed. We describe our proposed model's architecture and compare it with the comparison models in Section III. Section IV unfolds the performance evaluation metrics while V presents the results of each model in terms of accuracy, confusion matrices, classification reports, and model explainabilities. Section VI presents the concluding remarks and future scopes for the research.

II. Related Work

In this section, we discuss the current status of brain tumor diagnosis using DL detection and classification and define the problem statement accordingly.

In order to extract profound characteristics from MRI images of the brain, the study described in [22] utilized pre-existing models, such as Xception, NasNet Large, DenseNet121, and InceptionResNetV2. The dataset used in this investigation was Brain Tumor Detection 2020 (BR35H). There were 1500 normal photos and 1500 photos with tumors. The best classification performance was demonstrated by InceptionResNetV2 with a precision score of 99.68%. The authors in [24] tweaked the EfficientNet model. There were 3762 MR pictures in the collection. By using the aforementioned evaluation measures, the experiment calculated the number of correctly classified and incorrectly classified data and estimated the model's performance. 338 photos were properly classified as being free of tumors by the model, whereas five images were missed.

With the help of two datasets and a Gaussian Convolutional Neural Network (GCNN), [25] suggested a method for identifying different brain tumor types. The accuracy of the suggested method was 99.8% for the Medical University of Tianjin dataset and 97.14% for the "The Cancer Imaging Archive" (TCIA) dataset, respectively. GCNN's architecture consists of four convolutional layers that produced no overfitting or underfitting. The researchers in [26] suggested a hybrid encoder and decoder architecture (Hybrid-DANIet) that makes use of a number of modules integrated into a base encoder and decoder architecture. Numerous double convolutional blocks (DCBs) are incorporated in both the encoder and decoder of this U-shaped encoder-decoder design. They used the publicly accessible datasets BraTS 2017 (285 pictures) and BraTS 2018. (228 images). The authors in [27] proposed a three-step pre-processing strategy including removing confusing objects, denoising MRI images, and histogram equalization, and proposed an effective deep diagnosis system. The suggested Deep Convolutional neural network (DCNN) model achieved AUC scores of 94–96% for each of the classes on the dataset's 3394 pictures.

In [28], the authors tried to classify the images using convolutionary dictionary learning, which integrated a CNN architecture utilized for investigating distinctive information. In the CNN structure, the last layer contains the AlexNet and softmax classifiers. KNN graphs were used to make local constraints on atoms. Additionally, the authors examined the parameter sensitivity of convolutional dictionary learning with local constraint (CDLLC). Datasets from Cheng and REMBRANDT were retrieved from the public domain. The Cheng dataset has 3064 images with 3 classes. The REMBRANDT dataset contains 130 patient images which contain astrocytoma, oligodendroglioma, glioblastoma multiforme, plus other unclassified tumor types. The authors used the entire Cheng dataset and 1000 images from the REMBRANDT dataset for simulations, where the average value of 10 experiments is recorded to determine the performance. The authors presented an automatic method for segmenting tumors in MRI images that treated every image as a classification problem [29]. The authors used the dataset provided by the Multimodal Brain Tumor Segmentation Challenges of MICCAI 2012 and 2013. The data contains 120 subjects with gliomas, 55 from real patients, and 65 from synthetic data. The images have ground truth data associated with them and 80 images are used as training data. Using the learned softmax classification improved the accuracy of the proposed method, suggesting that the approach is more effective when more precise measurements are used.

The authors in [30] used a unique method of classification

where they broke the image into small segments and trained the model on that. This means that instead of a large model, they have small segmented data to work with, which reduces computation time. The idea was to decrease computing time and overcome the overfitting problem in cascade DL models. A Distance Wise Attention (DWA) algorithm was also used in this study. The paper used the BRATS 2018 dataset which contains 75 cases with Low-Grade Glioma (LGG) and 210 cases with High Grade Glioma (HGG). The training data is 80%, whereas validation and experimental data are 10% respectively. The accuracy obtained was dice scores of 0.9203, 0.9113, and 0.8726 for identifying the whole tumor, enhancing the tumor, and tumor core respectively. The experimental results were obtained using HAUSDORFF99, Dice similarity, and Sensitivity. In [31], the authors used two hybrid algorithms, ResNet-18 with support vector machines (SVMs) and GoogleNet with SVMs. In their study, they used a publicly available Figshare dataset that contains 3064 images of brain MRIs of 233 patients. Images showed Meningioma, Pituitary, and Glioma types of brain tumors. Through data augmentation, the authors increased the dataset 5 times, resulting in a total sample size of $15,320$ images. For ResNet-18, GoogLeNet, ResNet-18 with SVM, and GoogLeNet + SVM, the accuracy was 97.8%, 97.4%, 98.0%, and 97.6%, respectively. By breaking down the classification problem into a small one versus all subproblems, the authors aimed to classify tumors.

A random forest (RF) classifier was used instead of a single classifier in [32]. In a two-class fashion, each classifier predicted one tumor class. For this classification, the authors trained N distinct binary classifiers. The dataset was then post-processed to find overlapping predictions. According to the authors, the accuracy obtained with the whole tumor was 0.8999, with the tumor core was 0.7945, and with the active tumor was 0.7815.

Ferdous et al. in [33] examined different ML approaches to find the best-suited method for extracting and classifying brain tumors. To extract tumor features, the authors used a custom 10-layer CNN model. The model was tested on different ML techniques such as SVM, K-nearest neighbor (KNN), RF, Gaussian naive Bayes (G-NB), decision trees (DT), logistic regression (LR), and linear discrimination (LD). In total, 3064 patients were included in the dataset, with 994 axial, 1045 sagittal, and 1025 coronal groups, created by Cheng. Latif et al. in [34] proposed a system for dividing the MRI images into small blocks, where The discrete wavelet transform (DWT) method is utilized to extract the features in each block. Different types of tumors were classified using this method. DWT has also been used in watermarking techniques on medical images [35] In four different cases, the author used the BraTS benchmarked dataset provided by MICCAI with LGG and HGG images. Each case's ground truth values are also included in the dataset. RF trees, nearest neighbors, radial basis functions (RBFs), naive Bayes, and multi-layer perceptrons (MLPs) were used to classify the images. For HGG and LGG, the accuracy of RF and DWT was 89.75% and 86.87% respectively.

A hybrid GA-SVM method was used by the authors in [36] to select and classify the features. Data was collected from 55 patients with an overall value of 428. In the same way that genes are passed from generation to generation, genetic algorithms do the same. As the model trains, the fitness value increases, increasing accuracy. Sometimes weaker solutions are passed on, but this is helpful in recombination. GA-SVM provides more accurate results than other previous methods, according to the authors. In class 1 (AS), class 2 (GBM), class 3 (MEN), class 4 (MED), and class 5 (MET), the results were 89.8%, 83.3%, 94.5%, 96%, and 97.1%, respectively. With slight modifications, the authors in [37] proposed the ResNet50 algorithm for extracting robust features and learning the structure of MR images. Three linear modules have been added to the fully connected layer, two Leaky Relu modules have been added, two dropout modules have been added, and a softmax classification has been added to distinguish tumor types. Figshare contains 3064 MRI images of 233 patients with different types of brain tumors. As well as increasing the slice to 24512, the data were augmented. Maximum accuracy of 98.67% was obtained. With the help of a real-life dataset containing five classes of malignant tumors, the authors were able to differentiate between types of malignant tumors.

Vidyarthi et al. in [38] used a new cumulative variance method (CVM) to extract features from the data. KNN, mSVM, and NN were used to train and test the extracted features. Using the NN classifier, the proposed algorithm produced an accuracy of 95.86%. An encoder-decoder approach with three layers is used in [39] to learn diversified features. Dice loss and focal loss functions were combined. For training and validation, the authors used an RF regressor to train the features on shape, volumetric, and age which reach an accuracy of 56.7% and 51.7%, respectively. The authors in [40] used ensemble models, a similar approach to our proposed model. Based on 2556 images, this approach achieved an accuracy of 97.305%.

Each of the above-discussed models features ML and DL algorithms with impressive accuracy results. These models have some shortcomings that are listed in table I. DL models often classify parts of the data that do not contain the desired part. AI models are black box models, so it is imperative to deploy tools that can help us determine whether the models are predicting only the things we want.

III. METHODOLOGY

We follow some prime steps for the detection and classification of brain tumors in our research. Initially, we resize the images and augmented the data to multiply the number of images as part of the preprocessing. We fed data to six TL models with layers tweaked to increase their efficacy, the ensemble of the three best models (IVX16), and three ViT-based models. Analyzing and comparing the results of these models is performed after running them. Additionally, we assess the classification efficacy of the TL models and IVX16 models using LIME, an Explainable AI tool.

A. Dataset

We use a multiclass brain tumor dataset containing four classes of data including, Pituitary tumors, Glioma tumors,

TABLE I
COMPARISON OF RELATED WORK BRAIN TUMORS DETECTION AND CLASSIFICATION

Research	Contribution	Issues	AI models	Dataset size
[22]	Various DL techniques for the classification of brain tumor MRI images	Summary of Hyperparameters table seems unnecessary as all the variables are kept same for all the models	Xception, NasNet Large, DenseNet121, and InceptionResNetV2	3000
[24]	Tweaked the EfficientNet model where flattening, dropout, two FC layers, and a sigmoid classifier made up the EfficientNet-proposed B0's final layers	Accuracy and loss curves for the compared models could also be given in the paper	EfficientNet-B0, VGG16, InceptionV3, Xception, ResNet50 and InceptionResNetV2	3762
[25]	Four convolutional layers make up the GCNN's architecture, and they produced no overfitting or underfitting.	performance matrix could have been presented for other compared ML models	Gaussian Convolutional Neural Network (GCNN)	3064, 516
[26]	A hybrid encoder and decoder architecture that makes use of a number of modules integrated into a base encoder and decoder architecture	Larger datasets could have been considered	Hybrid-DANIet	285, 228
[27]	The suggested Deep Convolutional neural network (DCNN) model achieved AUC scores of 94–96%	AUCs of other compared models could have been shown along with the proposed model	VGG16, VGG19, CNN-SVM, and DCNN	3394
[28]	Classified the images using convolutionary dictionary learning	The authors did not implement any existing models to compare their results with the results of the proposed model	CDLLC	4064
[30]	Classification method where they broke the image into small segments and trained the model on	Authors could have employed any model validity tools to validate the performance of the proposed algorithm	DWA	285 cases
[29]	Automatic tumor segmentation method for MRI images	They used an online evaluation instead of checking with surgeons. Hence it is difficult to trust the results	local independent projection-based classification	120
[31]	Feature extraction using CNN-based model, ResNet-18 and GoogleNet and classification using SVM	More models could have been implemented to compare between their results	ResNet-18+SVM and GoogleNet+SVM	15320
[32]	Five RF classifiers instead of one to classify the tumors	A small size dataset is used	Random forest	274
[33]	A custom 10-layer CNN model for the tumor feature extraction and tested on different ML techniques	-	CNN-KNN	3064
[34]	Feature extraction using 3D DWT method for Blocks of MRI images	Limited number of cases for the type of modalities	Random Forest	18600
[36]	Genetic algorithm feature selection by sending features to the next generation in order to lead toward a fitness value	Not discussing the reasoning behind using different machines to take the data and very small size dataset	GA-SVM	428
[37]	Robust feature extraction and structure learning MRI images by changing the layers in ResNet50	Using only 15% of training dataset from a dataset of 24512	ResNet-50	24512
[38]	Features extraction from the data to derive crucial information from the data using a combination of CVM and a few other classifiers and	-	CVM+KNN, CVM+mSVM, CVM+NN	660
[39]	Feature learning using a 3D CNN with encoder-decoder approach	Used a very small dataset	RF	291
[40]	Ensembled-based model using RF, DT, and KNN	-	Hybrid Ensemble Classifier	2556
Our Proposed Model	An ensemble model based on the outputs of three transfer other learning models. Explainable AI in the form of LIME is used to confirm the validity of each of the models. Images are augmented making the total number from 3264 to 13056		IVX16 - an ensemble model based on the outputs of InceptionV3, Xcepton and VGG16	13056

Meningioma tumors, and no tumors. Our dataset contains 3264 images [41]. The dataset was partitioned into sets for training, testing, and validation, each with 80%, 10%, and 10% of data. With the ImageDataGenerator function, we virtually augmented the dataset by rescaling, shear range, zoom range, and horizontal flipping and increased the dataset four times from 3264 to 13056. A sample of data for each of the classes is shown in diagram 2.

B. Basic Architecture for the DL models

We used the same optimizer, loss function, and number of epochs for each of the six models: InceptionV3, VGG16, VGG19, ResNet50, InceptionResNetV2, and Xception. We maintain the dimensions of the input images at 224×224, weights from imagenet [42], and a batch size of 16. After making a sequential TL model, we add the convolutional layers followed by a flattening layer to make the individual TL model. After the convolutional layers of the TL models, the flatten layer flattens the multidimensional input tensors. The data is converted into a 1-D array before being passed to the

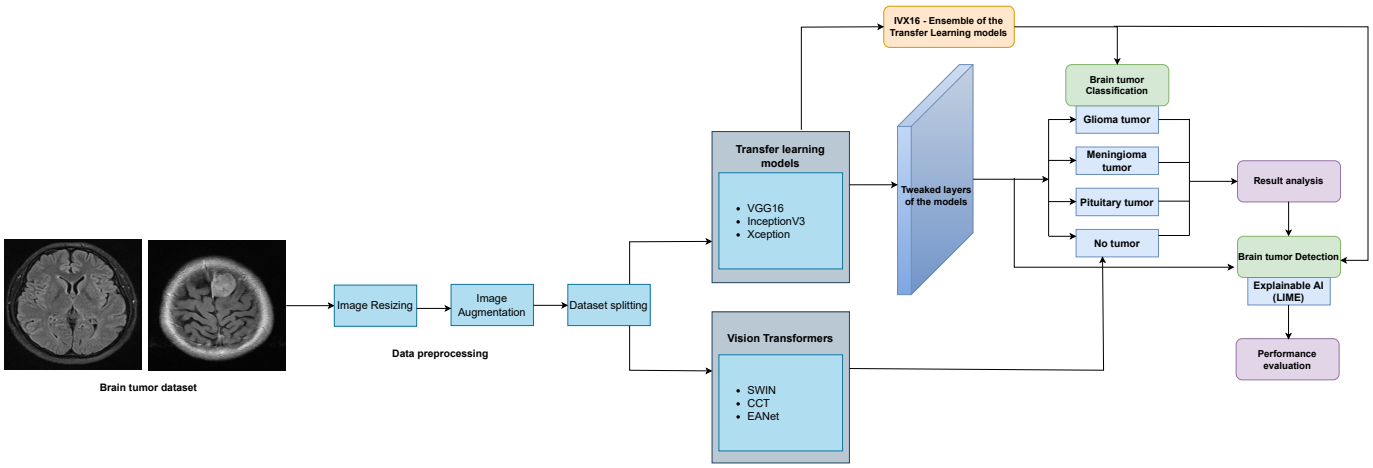

Fig. 1. Complete Block Diagram of procedures to classify brain tumors

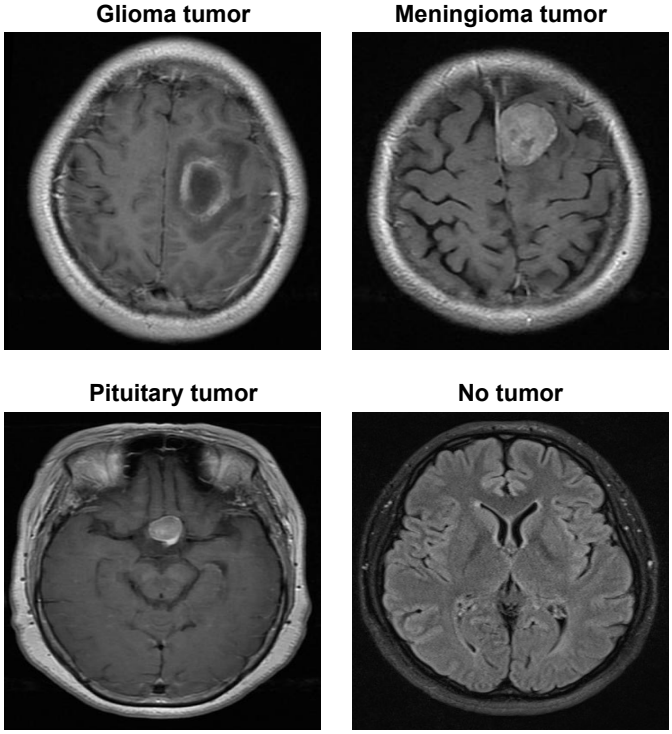

Fig. 2. Sample of MRI images

next layer. Subsequently, a dropout layer with a rate of 50% is appended to the flatten layer. During each iteration of training, this dropout layer removes half of the neuron connections from the model. A dense layer consisting of 512 neurons is then added where the activation function used is relu while for the final layer, 4 neurons are used for classifying each of the 4 types of brain tumors and here the activation function used is softmax. Due to the multiclass classification problem, the final layer in the model uses the softmax activation function. Categorical crossentropy was used as a loss function for the TL models as indicated in equation 1 with Adam as the optimizer while the learning rate was assigned to be $1e-5$. According to Adam, it is capable of determining the learning rates of specific parameters as part of the adaptive learning rate technique.

$$l = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}). \tag{1}$$

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{2}$$

In (2) sigma is the softmax function. In a multi-class classifier, the input vector is represented by the z vector, and its exponential function is denoted by the term standard exponential function $z(i)$. The output vector, on the other hand, comprises C classes, where the exponential function of $z(j)$ signifies the standard exponential function.

We obtain accuracy curves, loss curves, classification reports, and confusion matrices for all models under consideration after training. Based on all the result metrics generated by the model, we selected InceptionV3, VGG16, and Xception as the three best models for ensembling. InceptionV3 is more efficient than its predecessors $v1$ and $v2$, is computationally less expensive, and uses auxiliary classifiers as regularizers [43]. In 2014, VGG16 was used to win ILSVR (ImageNet) competition for its ability to fit complex functions and produce excellent results. As for multiple classes and large datasets, Xception is very efficient and makes efficient use of model parameters [44]. We chose these six models for the unique features exhibited by each of the models which are provided below.

*1) **VGG16**:* With VGG16, we can analyze images more effectively than ever before. VGG16 architecture converts images to $[224 \times 224]$ before feeding them into the model. The convolutional layers are constructed using kernels of $[3 \times 3]$ dimensions, and they exhibit varying depths. The first two layers contain 64 depths, while the fourth and fifth ones have 128 depths. The seventh, eighth, and ninth layers possess 256 depths, whereas the eleventh, twelfth, and thirteenth ones have 512 depths, all with a stride of 1. The Max pooling layers in the third, sixth, tenth, fourteenth, and eighteenth layers feature a $[2 \times 2]$ dimension with a stride of 2. Upon flattening the output, it is allocated to 4096 hidden layer units, and a layer

TABLE II
PARAMETERS FOR THE TRANSFER LEARNING MODELS AND ENSEMBLE MODEL

Models	Total parameters	Trainable parameters	Non Trainable parameters
InceptionV3	48,019,748	47,985,316	34, 432
VGG16	27,562,308	27,562,308	0
Xception	72,244,268	72,189,740	54, 528
VGG19	32,872,004	32,872,004	0
ResNet50	32,872,004	32,872,004	0
Inception-ResNetV2	74,000,100	73,939,556	60,544
IVX16	147,826,324	147,737,364	88, 960

with a softmax activation function is added. In the architecture, there are convolution layers, while in VGG16, there are two 3×3 layers.

*2) **InceptionV3**:* InceptionV3 is the model by Google, which is a modified version of the Inception architecture [45]. In spite of the fact that the model comprises 42 layers, its computation cost is only 2.5 times higher than GoogleNet's. It won the first runners-up prize for image classification from the ILSVRC.

*3) **Xception**:* Xception uses depthwise separable convolutions as part of its deep CNN architecture.

The Inception architecture involves the initial compression of the input using $[1 \times 1]$ convolutions, followed by the application of a unique set of filters to each depth space, which is determined by the input spaces. On the other hand, Xception works differently as prior to the input space compression with $[1 \times 1]$ convolution, filters are applied to each depth map. Since Xception does not introduce any nonlinearity, we use it.

*4) **VGG19**:* As a result of its 19 layered architecture, VGG19 is used due to its effects. At the apex of the architecture, the VGG19 model has three $[3X3]$ Convolution layers, which is one more than VGG16's two $[3X3]$ Convolution layers. Unlike VGG16, the max-pooling layer in VGG19 directly follows the convolutional layers, which are responsible for altering the input size from one layer to the next. Additionally, the architecture includes two dense, fully connected layers, each having 4096 neurons, and an output layer featuring 13 neurons.

*5) **ResNet50**:* ResNet50 consists of a residual architecture where the layers are assembled as learning residual functions concerning the input layers. The authors in [46] claimed that this is one of the deepest architectures.

*6) **InceptionResNetV2**:* As an extension of the Inception family of models, InceptionResNetV2 integrates residual connections [47]. It has been demonstrated empirically that training Inception networks with residual connections substantially speeds up the training process.

Our models have been tweaked to suit our convenience, so the number of trainable and non-trainable parameters has shifted. Table II shows a number of parameters for the different models we used.

C. Proposed Ensemble model

For better results, we build an ensemble model based on the above models. Our ensemble model, IVX16, aggregates predictions from VGG16, InceptionV3, and Xception. We use these three models out of the six as these three produced the best results which are discussed in Section V. There is a significant difference between the IVX16 and the other models in terms of performance, especially in terms of the detection of tumors discussed in part C of Experimental results.

Because TL models are trained on pre-trained data, they may not adapt as well to fresh data and produce biased results. Since our IVX16 model uses group weights, it is more robust to changes in the data or model architecture, so overfitting issues can be easily addressed. It is often the case that ensemble models are more accurate than individual models or TL models when compared to both. As demonstrated in the paper's result section, integrating the predictions of various models can decrease individual model errors and increase overall accuracy.

Furthermore, ensemble models in general, including the one discussed in this paper, are diverse, flexible, and interpretable in addition to improving performance and being robust. A diverse set of predictions is likely to result from the IVX16 because it is composed of three different models that are trained with different architectures. When dealing with noisy or complicated data, this approach is particularly effective. Assembling can be done with a wide range of models, making it very flexible and allowing us to use different models for different datasets. Ensemble models can shed light on which features or models are crucial for making predictions for interpretability or feature selection. Explainable AI is used in this paper to explore the interpretability of the TL models and IVX16 discussed later. The visual architecture of the IVX16 model is displayed in Fig. 3.

D. ViT models

There have been many image classification models introduced over the years. ViTs were introduced recently in the realm of image classification that originally originated from Natural Language Processing (NLP). Given a very large dataset, ViTs produce significantly better results than traditional DL algorithms. On small datasets such as the one used in this paper, ViT fails to produce the same remarkable results as it does on large ones.

$$Attention(Q, K, V) = softmax(QK^T/\sqrt{d_k})V. \quad (3)$$

*1) **SWIN**:* Shifted Windows transformers (SWIN) [52] is a variant of the ViT model. In this case, shifted windows are used to compute hierarchical Swin Transformer representations. While allowing cross-window connections, the shifted window technique enhances efficiency by restricting self-attention processing to non-overlapping local windows. The computational complexity of this design is linear with respect to image size, and it can simulate data at many scales.

The overall SWIN transformer is unfolded in Fig. 4. To begin the process of analyzing an RGB input image, a module such as ViT is employed to split it into separate, non-overlapping patches, known as "tokens." Each token's feature is created by concatenating the RGB values of individual pixels, resulting in a raw-valued feature projected

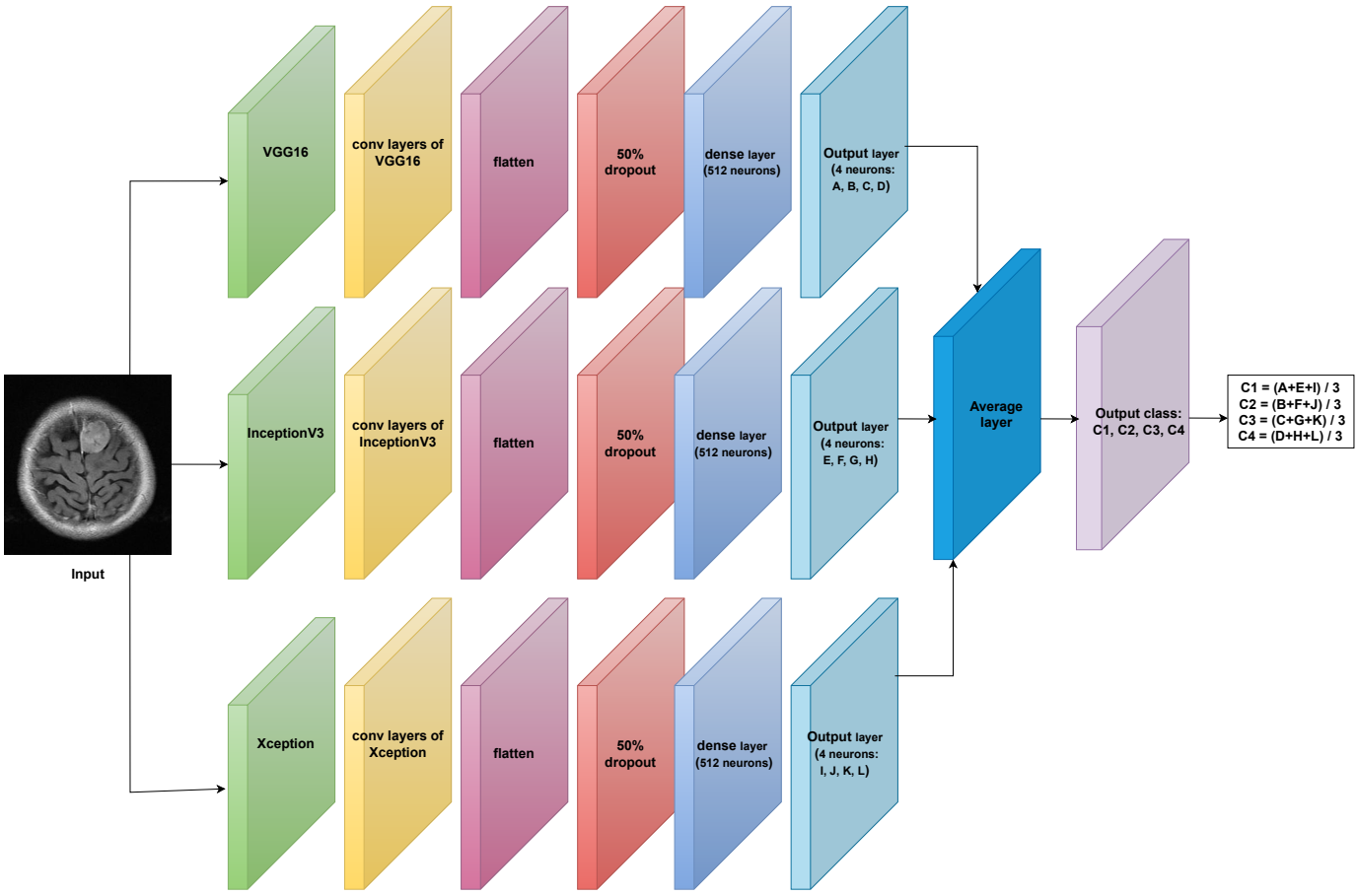

Fig. 3. Architecture of IVX16 model

to any dimension via a linear embedding layer. A series of Transformer blocks, which have modified self-attention computation, process these patch tokens. The linear embedding and Transformer blocks, together preserving the number of tokens (H/4 W/4), form "Stage 1."

As the network goes deeper, a reduction in the number of tokens is achieved by utilizing patch merging layers. The first patch merging layer performs a linear operation on the 4C-dimensional merged features obtained by concatenating the features of every adjacent set of 2×2 patches. This process results in a decrease of tokens by a multiple of $2 \times 2 = 4$, and the output size is set to 2C. After the application of Swin Transformer blocks for feature transformation, the resolution is kept constant at $H/8 \times H/8$. The first phase of patch merging and feature transformation is known as "Stage 2". The process is repeated twice, referred to as "Stage 3" and "Stage 4", respectively, with output resolutions of $H/16 \times H/16$ and $H/32 \times H/32$. In combination, these stages generate a hierarchical representation with feature map resolutions comparable to those seen in traditional convolutional networks. The entire architecture of the Swin transformer is shown in Fig. 4.

To create Swin Transformer, the standard multi-head self-attention (MSA) module in a Transformer block is substituted with a shifted window-based module, while the other layers remain unchanged.

*2) **CCT**:* With CCT, convolutional and transformer techniques are synergistically combined to enhance visual effects. Unlike traditional ViT models which utilize non-overlapping patches, CCT makes use of convolution when local information can be utilized more effectively [53]. Probabilistic depth tuning is used to address the issue of the vanishing gradient in the CCT model. Similar to dropout, this approach involves the arbitrary elimination of slices at varying depths in the model. After undergoing convolutional tokenization, the input data in CCT is encoded using a transformer. Sequence pooling MLP headers facilitate the possibility of multiple dissection varieties of brain tumor diseases. Fig. 5 unfolds the entire CCT operation.

*3) **EANet**:* EANet [54] bases its operation on two external, small, teachable, and shared memories, M_k and M_v. By deleting patches with redundant and unnecessary information, EANet increases performance and computational efficiency. In order to implement external attention, two linear layers and two normalization layers are used. EANet determines the attention between input pixels and an external memory unit as follows.

$$A = Norm(FM_k^T), \tag{4}$$

$$Fout = AM_v. \tag{5}$$

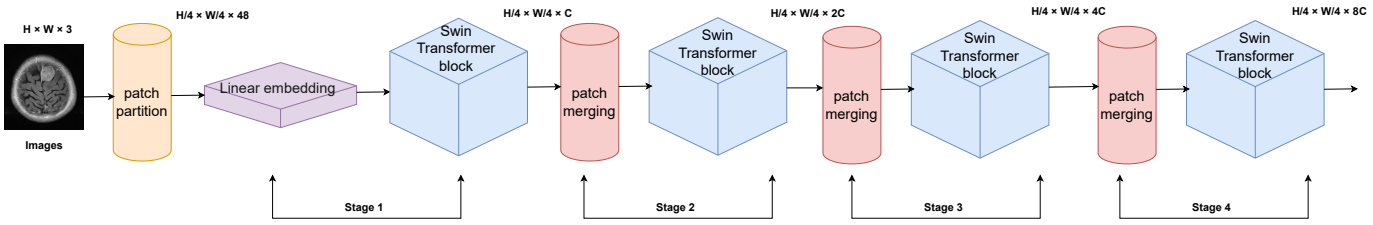

Fig. 4. Architecture of SWIN transformer model

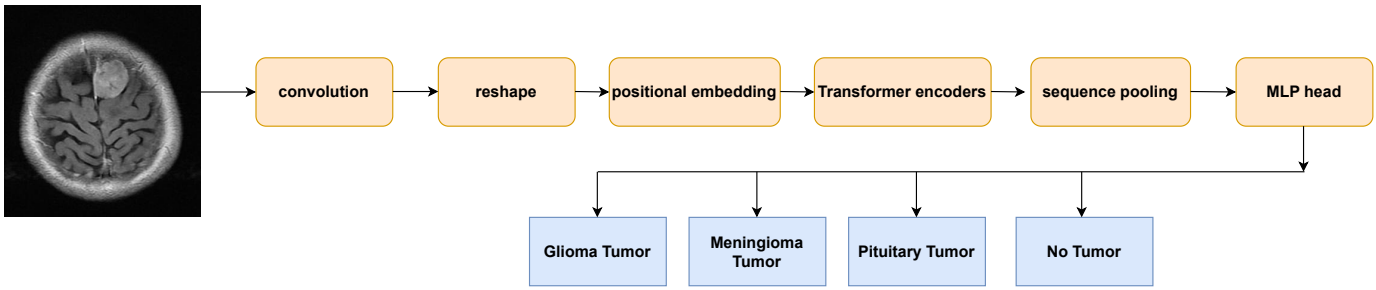

Fig. 5. Compact Convolutional Transformer architecture

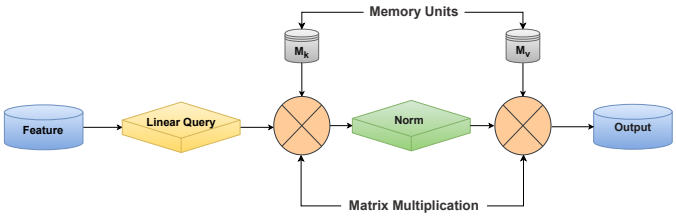

Fig. 6. External-attention for EANet model

Finally, attention to A's similarities updates M_v's input characteristics. The attention model diagram is displayed in Fig. 6.

IV. Performance Evaluation Metrics

Several metrics are used to evaluate the performance of each model. We provide accuracy curves, confusion matrices, classification reports, and Explainable AI.

An accuracy curve gives us an indication of a model's highest accuracy, and its smoothness shows the model's performance as a classifier. Having a smoother curve ensures that the model is a better classifier. The confusion matrix compares the actual label with the anticipated category label in a two-dimensional array. In this way, it is easier to keep track of how many images are correctly and incorrectly classified by the models.

Each model's effectiveness is measured by four metrics in our classification report. There are several metrics used to evaluate the model's performance, including Support, Precision, F1 score, and Recall. Based on the last rows of the confusion matrix, support refers to the number of instances in each class's actual answers.

The precision(P) refers to the proportion of correctly predicted results and the total number of positively classified observations. The following is a definition of precision.

$$P = \frac{T_p}{T_p + F_p}, \tag{6}$$

The recall(R) is calculated by dividing the number of predicted results by all the evaluations of the original class.

$$R = \frac{T_p}{T_p + F_n}, \tag{7}$$

The *F1*-score(F1) calculates a single score by averaging precision and recall.

$$F1 = \frac{2 \times precision \times recall}{precision + recall}, \tag{8}$$

In the above set of equations,

- T_p = True Positive
- F_p = False Positive
- F_n = False Negative
- T_n = True Negative

In order to provide appropriate care to patients, we must also determine whether our models are correctly classifying the images provided, and whether our classifications can be trusted. Our models may classify parts of the images that are outside the region of interest. Our model must identify and classify the cancerous regions on the images, not the borders of the image frames, to identify and classify the specific types and locations of tumors. It is therefore crucial to visualize the model results before drawing any conclusions.

We use LIME [48] to visualize the tumor-affected regions. "L" stands for Local, which means the explanations are local, not global, so they can be expressed in terms of individual images, not the whole dataset. "I" stands for Interpretable, which means humans cannot understand the parameters of a neural network, such as its weights. Humans can, however, interpret the results produced by LIME. "M" stands for Model-Agnostic, which means that LIME can be used for any

ML models and datasets. In our case, we are dealing with image data. "E" stands for Explanations, which means LIME provides an understanding of the model's input to model's output prediction. LIME can be defined as follows.

$$explanation(x) = L(f, g, \Pi_x) + \Omega(g). \tag{9}$$

An explanation model for a given example x is denoted by the model g, which could be a linear regression (LR) model, capable of significantly reducing the loss L (e.g., mean squared error) and maintaining a low model complexity $\Omega(g)$ (preferably fewer features). The G family encompasses all possible LR models, among others, that serve as explanations. The proximity measure (x) determines the region surrounding x that is considered while providing the explanation. LIME, in essence, enhances the loss component. The user must choose the maximum amount of features the LR model may employ, for example, in order to define the complexity.

V. Experimental results

A. Ensemble model and comparison model results

Validation accuracy of the TL models InceptionV3, VGG16, Xception, ResNet50, VGG19, InceptionResNetV2, and the ensembles model IVX16 are represented in table III along with the accuracies of some SOTA models implemented in related research papers. The best three performing models in terms of validation accuracy are InceptionV3, VGG16, and Xception has been used to develop the ensemble model IVX16. Our model IVX16 produces a peak accuracy of 96.94%.

TABLE III
Model Accuracies

Models	**Accuracy**
InceptionV3	95.72
VGG16	95.11
Xception	94.50
IVX16	96.94
ResNet50	93.88
VGG19	94.19
InceptionResNetV2	93.58

In training and validation, IVX16 performs better than the other individual TL models. Using an ensemble model simplifies things when using Explainable AI tools whose visualizations are heavily dependent on the layers of the model since it incorporates features from the base models used to make it. Fig. 7 demonstrates the comparison of each model in terms of accuracy curves. As compared to the other models, the IVX16 achieves greater peak accuracy and has greater average accuracy. Other TL models also perform well. As the number of epochs increases, InceptionV3's validation accuracy fluctuates less and less, while the difference between its training and validation accuracy decreases. VGG16 and VGG19 also display good performance as per their respective accuracy curves, however, VGG16 shows better average validation accuracy than VGG19. For the first 10 epochs Xception shows random accuracy values with its fluctuating validation curve, after that, we see a pretty steady curve and good improvement in validation accuracy. ResNet50 shows slight overfitting but shows decent accuracy throughout the whole training period while InceptionResNetV2 shows an accuracy drop between 10-15 epochs but recovers afterward and produces a decent accuracy curve.

Fig. 8 shows the confusion matrices for each of the models. The confusion matrices show that all the models struggle to classify Glioma tumors. However, IVX16 produces better results when classifying Glioma tumors than the other models. IVX16 gives a score of 0.49 meaning it correctly classifies around 49% of the total Glioma images. Meningioma tumors are classified using all six TL models and our ensemble model. While classifying Pituitary Tumors, IVX16 also performs better than most TL models. The diagram shows that all the models produce almost a 100% success rate when inspected with images that do not contain tumors.

In the investigation of the models in the classification report in table IV, we see IVX16 outclasses all the other TL models by obtaining an F1 score of 0.61 for classifying Glioma tumors. IVX16 also scores the highest precision score for Meningioma which is 0.73. Only Xception has a closer precision value for Meningioma, producing a value of 0.71. In terms of the F1 scores for the Meningioma class, IVX16 again outperforms all the other models scoring a value of 0.82 which is higher than all other TL models. For the analysis of Pituitary tumors, IVX16 along with InceptionV3 and VGG19 shows a perfect score of 1.00 in terms of precision. In terms of F1 score, VGG16 exhibits the highest value of 0.88 whereas IVX16 produces the second highest value of 0.85. For scrutinizing images with no tumors, we discover from the classification report table that all the models struggle to extract a decent score for precision, this is due to the class having a lower number of data along training, testing, and validation sets compared to other classes. For F1 scores, IVX16 and VGG16 produce the highest scores of 0.75 and 0.77 respectively. In the macro average column, IVX16 dominates over the other TL models with a value of 0.76 in terms of F1 score. The precision values for all the models stay very similar for the macro average. In the weighted average column, IVX16 outdo all the TL models in terms of both recall and F1 score. Since IVX16 produces better quantitative results for most of the metrics it is fair to say that IVX16 is the best classifier among all the models.

B. ViT models results

From both Fig. 9, we see clear signs of overfitting for each of the ViT models. Unlike the TL and IVX16 models which ran on 30 epochs, we run the ViT models on 100 epochs but still fail to observe any improvement in the results. In the confusion matrix 10 we see the models fail to classify the images in all of the classes and produce very poor values. No tumor class is an exception here, in this class, the ViT models show a very high classification rate. In the classification report of the ViT models in Table V, we once again see poor values across all the classification metrics for each of the models. The highest value observed is 0.80 which is the precision value for the SWIN transformer on the Glioma tumor, we additionally observe some decent values like 0.74 and 0.76 which are

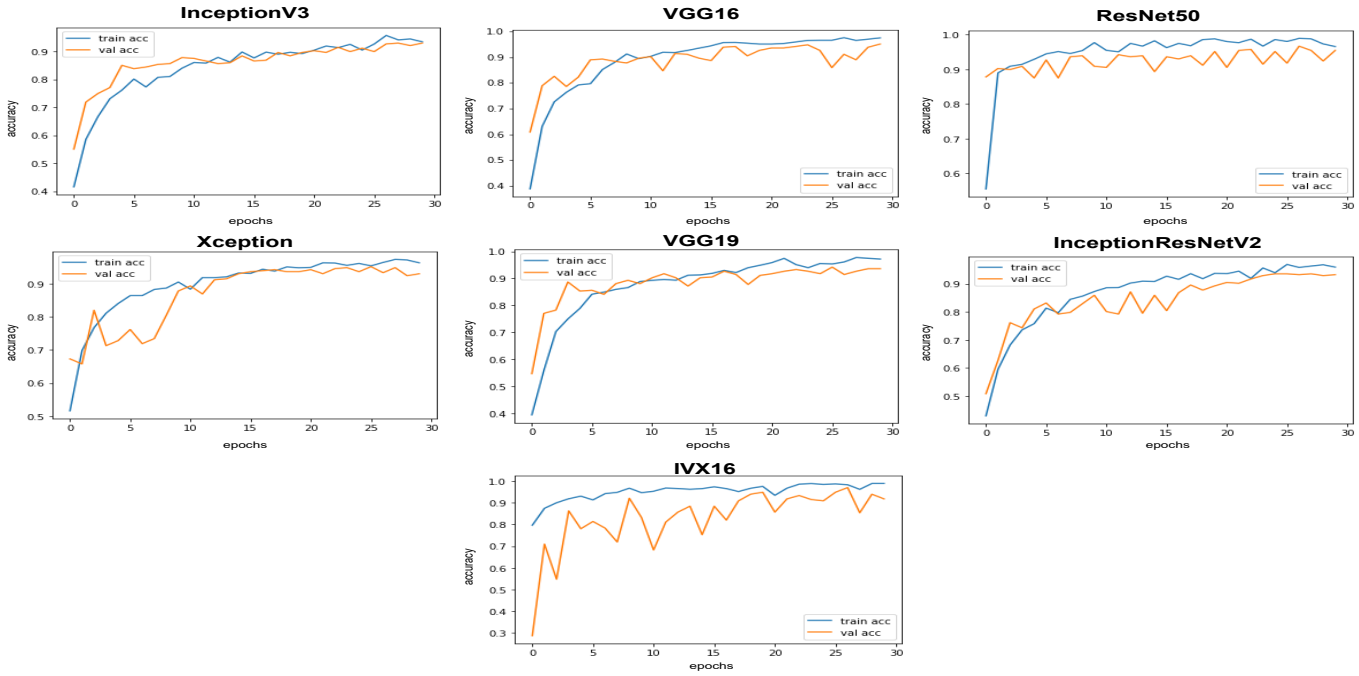

Fig. 7. Accuracy curves for Transfer learning and ensemble model

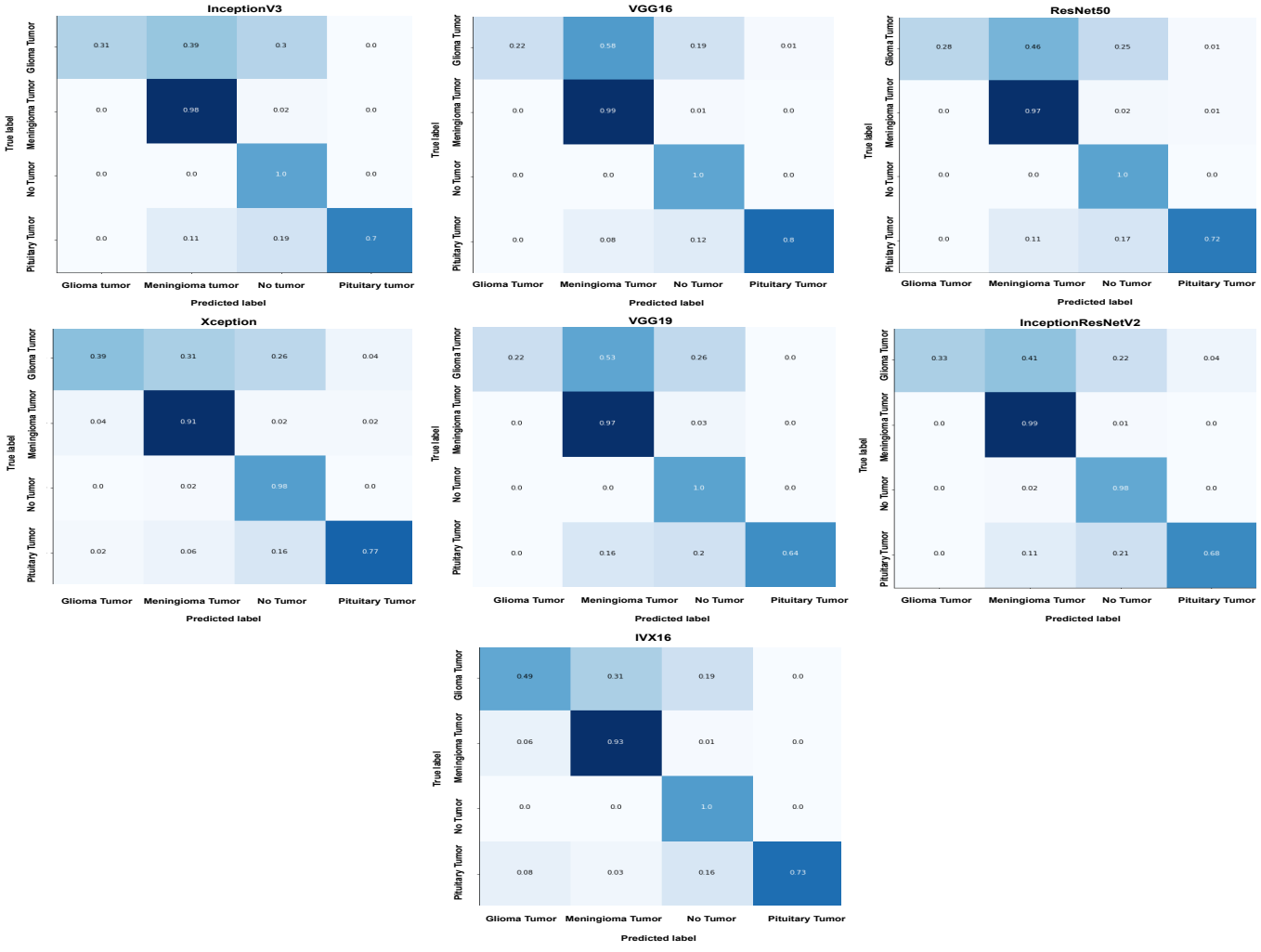

Fig. 8. Comparison of Confusion matrices for the Ensemble and Transfer Learning models

TABLE IV
CLASSIFICATION REPORT FOR TRANSFER LEARNING AND ENSEMBLE MODEL

Models	Metrics	Glioma	Meningioma	Pituitary	No tumor	Macro average	Weighted average
InceptionV3	Precision	1.00	0.67	1.00	0.52	0.80	0.83
	Recall	0.31	0.98	0.70	1.00	0.75	0.72
	F1 score	0.48	0.79	0.82	0.68	0.69	0.69
	Support	93	94	90	50	327	327
VGG16	Precision	1.00	0.60	0.99	0.62	0.80	0.83
	Recall	0.22	0.99	0.80	1.00	0.75	0.72
	F1 score	0.35	0.75	0.88	0.77	0.69	0.68
	Support	93	94	90	50	327	327
ResNet50	Precision	1.00	0.63	0.97	0.56	0.79	0.82
	Recall	0.28	0.97	0.72	1.00	0.74	0.71
	F1 score	0.44	0.76	0.83	0.71	0.69	0.68
	Support	93	94	90	50	327	327
Xception	Precision	0.86	0.71	0.92	0.55	0.76	0.79
	Recall	0.39	0.91	0.77	0.98	0.76	0.73
	F1 score	0.53	0.80	0.84	0.71	0.72	0.72
	Support	93	94	90	50	327	327
VGG19	Precision	1.00	0.59	1.00	0.53	0.78	0.81
	Recall	0.22	0.97	0.64	1.00	0.71	0.67
	F1 score	0.35	0.73	0.78	0.69	0.64	0.63
	Support	93	94	90	50	327	327
InceptionResNetV2	Precision	1.00	0.65	0.94	0.55	0.79	0.82
	Recall	0.33	0.99	0.68	0.98	0.75	0.72
	F1 score	0.50	0.79	0.79	0.71	0.70	0.69
	Support	93	94	90	50	327	327
IVX16	Precision	0.78	0.73	1.00	0.60	0.78	0.80
	Recall	0.49	0.93	0.73	1.00	0.79	0.76
	F1 score	0.61	0.82	0.85	0.75	0.76	0.75
	Support	93	94	90	50	327	327

precision values for the Glioma tumor and Meningioma tumor by CCT. But throughout the entire table, we see a continuous display of poor values of classification across different metrics.

Data from this dataset is very sparse, so this is to be expected. ViT models need a huge bulk of data to produce significant results but they fail to replicate such remarkable results for datasets like the comparatively small one we are using. Since these models fail to produce any distinguishable results on the classification of brain tumors we do not test the validity of the models in terms of the detection of tumors using XAI.

C. TL and Ensemble Model Explainability by LIME

We generated three rows of images using LIME for three different types of tumors that we classified based on the best three TL models and our ensemble-based IVX16 model. To explore the performance of each model, we randomly select images from each of the three classes.

1) Pituitary tumors: For the same Pituitary tumor image, the first row of Fig. 11 shows the features of the LIME output images produced by each model. As seen in InceptionV3, the model classifies parts of the images that are not affected, while the original image only has one small area affected. Pituitary tumors are incorrectly visualized and located in VGG16. Xception classifies a region near the tumor area somewhat. The proposed IVX16 model does not necessarily provide a completely accurate visualization, but one of the yellow patches corresponds to the location of the tumor in the original image, which makes it the most accurate model for classifying and identifying the tumor. While InceptionV3 and VGG16 produce good accuracy results, they fail to detect tumors because they may have classified parts of the image outside the region of interest during training. TL models fail to detect the correct tumorous region due to their less robust nature compared to IVX16 models. Because IVX16 uses group weights, it is more robust and capable of solving complex patterns and detecting tumorous regions that cannot be detected by TL models.

2) Glioma tumors: Next, we generate LIME images of a randomly selected image with a glioma tumor. Multiple green patches appear on regions of the brain in this row image generated by LIME on InceptionV3. Furthermore, InceptionV3 falsely classifies a large chunk of the lower right part of the image as tumorous on the upper part of the brain. It engulfs the region of the tumor but also classifies parts of images that are not tumorous. Both VGG16 and Xception fail to locate the tumor and classify healthy brain tissue as tumorous as they do not have the group weights that IVX16 utilizes. The ensemble model IVX16 produces three different green patches, out of which one completely encloses the tumorous region, making IVX16 the best classifier.

3) Meningioma Tumors: In the final row, LIME visualizations are applied to a Meningioma tumor image. InceptionV3 produces a gigantic patch that fails to detect the tumor. VGG16 produces giant patches that classify more than 50% of the image to be tumorous. One of the patches captures the original region of the tumor but again additionally classifies some extra portion of the image having no tumors. Xception produces two patches and one of the patches successfully encloses the tumor located. IVX16 produces gigantic patches which partly consume the tumorous region but also consume a large region

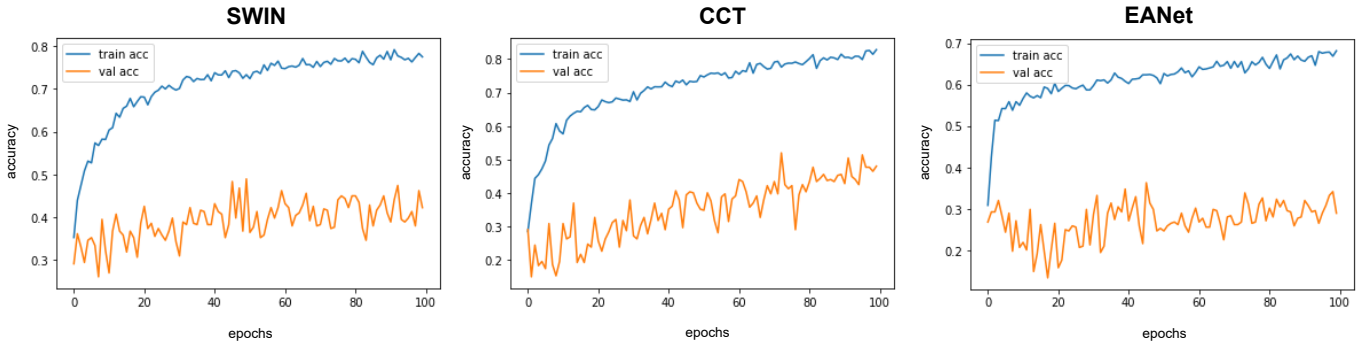

Fig. 9. Accuracy curves for ViT models

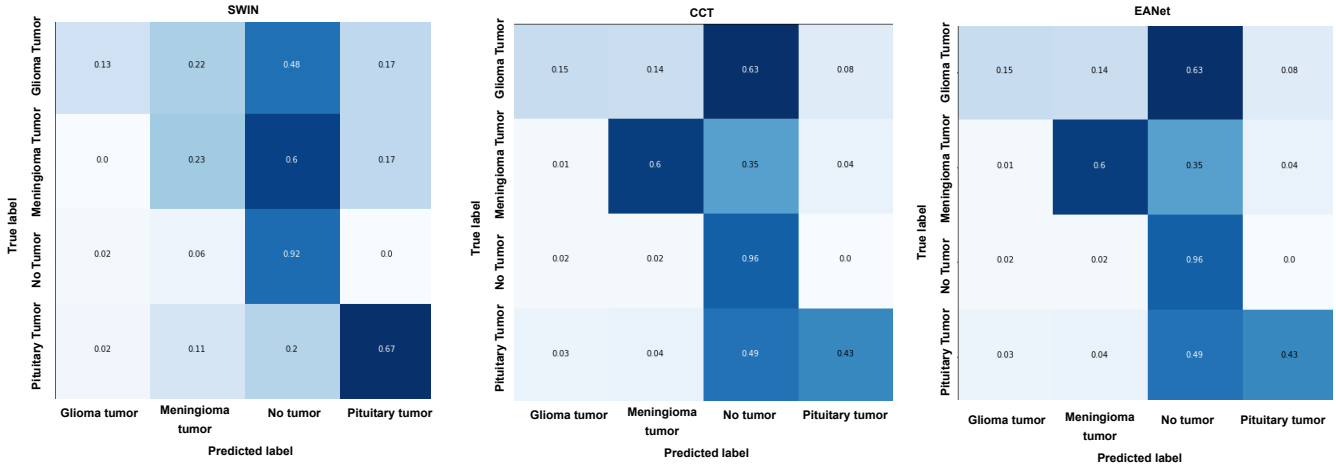

Fig. 10. Comparison of Confusion matrices for ViT models

TABLE V
CLASSIFICATION REPORT FOR VIT MODELS

Models	Metrics	Glioma	Meningioma	Pituitary	No tumor	Macro average	Weighted average
SWIN	Precision	0.80	0.40	0.65	0.28	0.53	0.56
	Recall	0.13	0.23	0.67	0.92	0.49	0.43
	F1 score	0.22	0.30	0.66	0.43	0.40	0.39
	Support	93	94	90	50	327	327
CCT	Precision	0.74	0.76	0.78	0.26	0.63	0.68
	Recall	0.15	0.60	0.43	0.96	0.53	0.48
	F1 score	0.25	0.67	0.56	0.41	0.47	0.48
	Support	93	94	90	50	327	327
EANet	Precision	0.56	0.23	0.61	0.18	0.39	0.42
	Recall	0.05	0.12	0.49	0.70	0.34	0.29
	F1 score	0.10	0.15	0.54	0.28	0.27	0.27
	Support	93	94	90	50	327	327

without any tumors.

VI. CONCLUSION

To classify tumors into different classes, we investigated several models in this paper. Then, we proposed a novel ensemble and TL-based model for classifying brain tumors, IVX16, and compared its output with established TL algorithms VGG16, VGG19, InceptionV3, ResNet50, Inception-ResNetV2, and Xception. Our proposed model was based on the ensembling of the best three performing models which are VGG16, InceptionV3, and Xception. By assembling the models, we were able to make a more robust model, since the ensemble models are more likely to avoid overfitting issues, and the complexity of the models enables it to deal with complex patterns in the images. It was found that some of the results produced by our proposed model were better and comparable to those produced by the above SOTA algorithms. Moreover, we compared the outputs of ViT models with those of traditional TL models and ensemble models. Additionally, we validated the performance of each of the TL models and IVX16 using LIME, an explainable AI tool. Our model was more accurate than others discussed in LIME analysis when

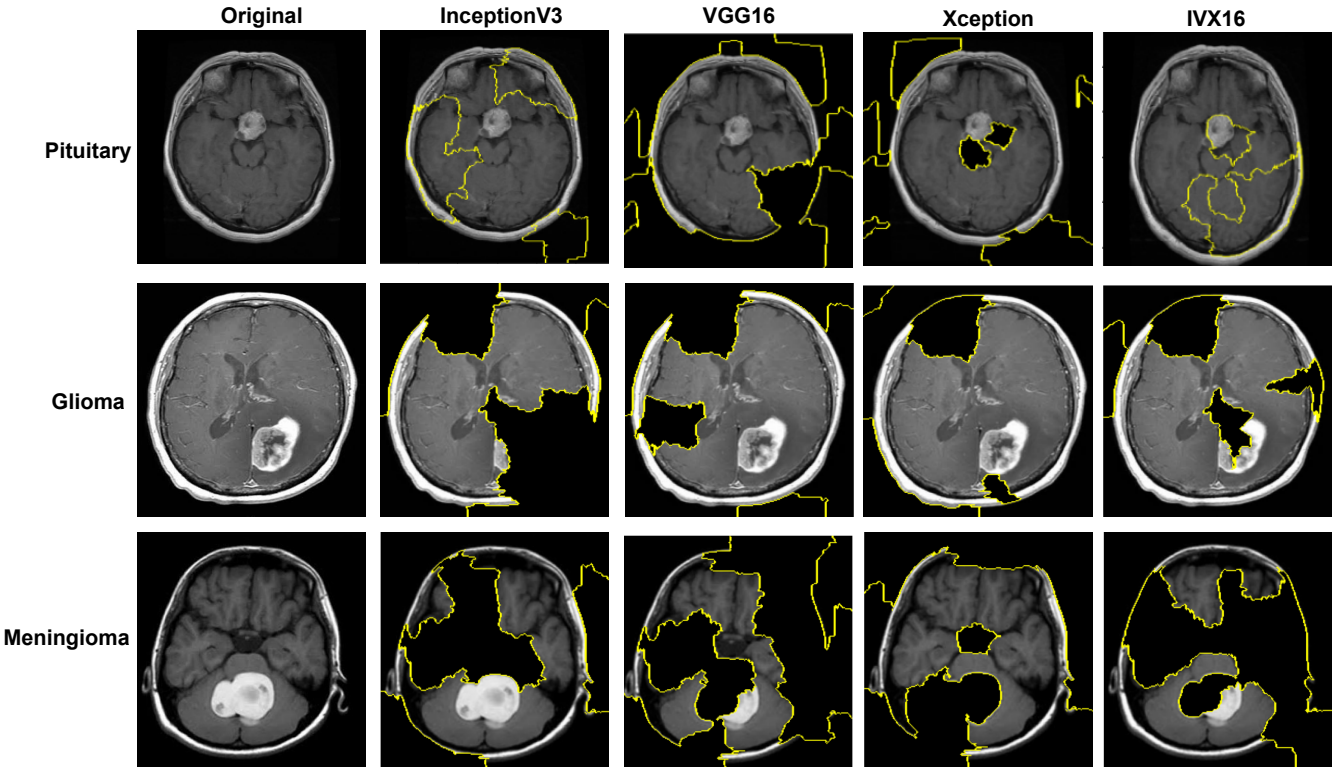

Fig. 11. Pituitary, Glioma, and Meningioma tumor region images generated using LIME

it came to detecting tumor regions in images.

ACKNOWLEDGMENT

This work was supported by the Korean Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries(IPET) through Digital Breeding Transformation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (322063-03-1-SB010).

REFERENCES

[1] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran, and M. Shoaib, "A deep learning model based on concatenation approach for the diagnosis of brain tumor," IEEE Access, vol. 8, pp. 55 135–55 144, March 2020.

[2] A. Wulandari, R. Sigit, and M. M. Bachtiar, "Brain Tumor Segmentation to Calculate Percentage Tumor Using MRI," in Proc. of the International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), Bali, Indonesia, pp. 292–296, 2018.

[3] E. S. Chahal, A. Haritosh, A. Gupta, K. Gupta, and A. Sinha, "Deep Learning Model for Brain Tumor Segmentation & Analysis," in Proc. of the 3rd International Conference on Recent Developments in Control, Automation Power Engineering (RDCAPE), Noida, India, pp. 378–383, 2019.

[4] Wang, Chaoyue, Aurea B. Martins-Bach, Fidel Alfaro-Almagro, Gwenaëlle Douaud, Johannes C. Klein, Alberto Llera, Cristiana Fiscone et al. "Phenotypic and genetic associations of quantitative magnetic susceptibility in UK Biobank brain imaging." Nature Neuroscience: 1-14, May 2022.

[5] M. I. Sharif, J. P. Li, M. A. Khan, and M. A. Saleem, "Active deep neural network features selection for segmentation and recognition of brain tumors using mri images," Pattern Recognition Letters, vol. 129, pp. 181–189, November 2019.

[6] Liu, Zhihua, Lei Tong, Long Chen, Zheheng Jiang, Feixiang Zhou, Qianni Zhang, Xiangrong Zhang, Yaochu Jin, and Huiyu Zhou. "Deep learning based brain tumor segmentation: a survey." Complex and Intelligent Systems: 1-26, July 2020.

[7] M. A. Ottom, H. A. Rahman, and I. D. Dinov, "Znet: Deep learning approach for 2d mri brain tumor segmentation," IEEE Journal of Translational Engineering in Health and Medicine, vol. 10, pp. 1–8, May 2022.

[8] H. S. Abdulbaqi, K. N. Mutter, M. Z. M. Jafri, and Z. A. Al-Khafaji, "Estimation of brain tumour volume using expanded computed tomography scan images," in Proc. of the 23rd Iranian Conference on Biomedical Engineering and 2016 1st International Iranian Conference on Biomedical Engineering (ICBME), Tehran, Iran, pp. 117–121, 2016.

[9] R. Sethi, M. Mehrotra, and D. Sethi, "Deep Learning based Diagnosis Recommendation for COVID-19 using Chest X-Rays Images," in Proc. of the Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, pp. 1–4, 2020.

[10] y. p. Sammy V. Militante and Nanette V. Dionisio and Brandon G. Sibbaluca, "Pneumonia detection through adaptive deep learning models of convolutional neural networks.", 11th IEEE Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia, pp. 88-93, 2020.

[11] S. Hussein, P. Kandel, C. Bolan, M. Wallace, and U. Bagci, "Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches," IEEE Transactions on Medical Imaging, vol. 38, no. 8, pp. 1777–1787, Aug. 2019.

[12] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), Brooklyn, NY, USA, pp. 294–297, 2015.

[13] Z. Lv, L. Qiao and A. K. Singh, "Advanced Machine Learning on Cognitive Computing for Human Behavior Analysis," in IEEE Transactions on Computational Social Systems, vol. 8, no. 5, pp. 1194-1202, Oct. 2021.

[14] Ali, T. M., Nawaz, A., Ur Rehman, A., Ahmad, R. Z., Javed, A. R., Gadekallu, T. R., Chen, C., and Wu, C., "A Sequential Machine Learning-cum-Attention Mechanism for Effective Segmentation of Brain Tumor," in Frontiers in Oncology, vol. 12, June 2022.

[15] B. M, "Automatic Segmenting Technique of Brain Tumors with Convolutional Neural Networks in MRI Images," In. Proc. of the 6th International Conference on Inventive Computation Technologies (ICICT), pp. 759-764, Coimbatore, India, 2021.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Proc. of the 25th

International Conference on Neural Information Processing Systems - Volume 1, ser. NIPS'12. Red Hook, NY, USA: Curran Associates Inc., p. 1097–1105, Jan. 2012.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[18] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1207–1216, Feb. 2016.

[19] C. Ding and D. Tao, "Robust Face Recognition via Multimodal Deep Face Representation," IEEE Transactions on Multimedia, vol. 17, no. 11, pp. 2049–2058, Nov. 2015.

[20] Q. Wang, F. Liu, G. Wan, and Y. Chen, "Inference of brain states under anesthesia with meta learning based deep learning models," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 1081–1091, May 2022.

[21] Y. xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, "Car detection from low-altitude UAV imagery with the faster R-CNN," Journal of Advanced Transportation, vol. 2017, pp. 1–10, Aug. 2017.

[22] S. Asif, W. Yi, Q. U. Ain, J. Hou, T. Yi and J. Si, "Improving Effectiveness of Different Deep Transfer Learning-Based Models for Detecting Brain Tumors From MR Images," in IEEE Access, vol. 10, pp. 34716-34730, 2022.

[23] M. Gouskir, M. A. Zyad and M. Boutalline, "Automatic Analysis of Brain Tumor from Magnetic Resonance Images based on Geometric Median Shift," 2020 IEEE 6th International Conference on Optimization and Applications (ICOA), pp. 1-7, Beni Mellal, Morocco, 2020.

[24] H. A. Shah, F. Saeed, S. Yun, J. -H. Park, A. Paul and J. -M. Kang, "A Robust Approach for Brain Tumor Detection in Magnetic Resonance Images Using Finetuned EfficientNet," in IEEE Access, vol. 10, pp. 65426-65438, 2022.

[25] M. Rizwan, A. Shabbir, A. R. Javed, M. Shabbir, T. Baker and D. Al-Jumeily Obe, "Brain Tumor and Glioma Grade Classification Using Gaussian Convolutional Neural Network," in IEEE Access, vol. 10, pp. 29731-29740, 2022.

[26] N. Ilyas, Y. Song, A. Raja and B. Lee, "Hybrid-DANet: An Encoder-Decoder Based Hybrid Weights Alignment With Multi-Dilated Attention Network for Automatic Brain Tumor Segmentation," in IEEE Access, vol. 10, pp. 122658-122669, 2022.

[27] A. S. Musallam, A. S. Sherif and M. K. Hussein, "A New Convolutional Neural Network Architecture for Automatic Detection of Brain Tumors in Magnetic Resonance Imaging Images," in IEEE Access, vol. 10, pp. 2775-2782, 2022.

[28] Gu, X., Shen, Z., Xue, J., Fan, Y., and Ni, T. Brain Tumor MR Image Classification Using Convolutional Dictionary Learning With Local Constraint. Frontiers in neuroscience, vol. 15, pp 679847, May 2021.

[29] Huang, Meiyan Yang, Wei Wu, Yao Jiang, Jun Chen, Wufan and Feng, Qianjin, "Brain Tumor Segmentation Based on Local Independent Projection-Based Classification", IEEE transactions on bio-medical engineering. 61. 2633-2645, Oct. 2014.

[30] Ranjbarzadeh, R., Bagherian Kasgari, A., Jafarzadeh Ghoushchi, S. et al., "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images", Sci Rep 11, 10930, May 2021.

[31] H. Kibriya, M. Masood, M. Nawaz, R. Rafique, and S. Rehman, "Multiclass Brain Tumor Classification Using Convolutional Neural Network and Support Vector Machine," in Proc. of the Mohammad Ali Jinnah University International Conference on Computing (MAJICC), Karachi, Pakistan, pp. 1–4, 2021.

[32] M. T. El-Melegy and K. M. A. El-Magd, "A Multiple Classifiers System for Automatic Multimodal Brain Tumor Segmentation," in Proc. of the 15th International Computer Engineering Conference (ICENCO), Cairo, Egypt, pp. 110–114, 2019.

[33] G. J. Ferdous, K. A. Sathi, and M. A. Hossain, "Application of Hybrid Classifier for Multi-class Classification of MRI Brain Tumor Images," in Proc. of the 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Dhaka, Bangladesh, pp. 1–6, 2021.

[34] G. Latif, M. M. Butt, A. H. Khan, O. Butt, and D. N. F. A. Iskandar, "Multiclass brain Glioma tumor classification using block-based 3D Wavelet features of MR images," in Proc. of the 4th International Conference on Electrical and Electronic Engineering (ICEEE), Ankara, Turkey, pp. 333–337, 2017.

[35] Sharma, A., Singh, A. K., and Ghrera, S. Secure Hybrid Robust Watermarking Technique for Medical Images. Procedia Computer Science, vol. 70, pp 778-784, Nov. 2015. https://doi.org/10.1016/j.procs.2015.10.117

[36] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, and C. K. Ahuja, "Multiclass Brain Tumor Classification Using GA-SVM," in Proc. of the Developments in E-systems Engineering, pp. 182–187, Dubai, United Arab Emirates, 2011.

[37] D. S, L. Padma Suresh, and A. John, "A Deep Transfer Learning framework for Multi Class Brain Tumor Classification using MRI," in Proc. of the 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, pp. 283– 290, 2020.

[38] A. Vidyarthi, R. Agarwal, D. Gupta, R. Sharma, D. Draheim, and P. Tiwari, "Machine Learning Assisted Methodology for Multiclass Classification of Malignant Brain Tumors," IEEE Access, vol. 10, pp. 50 624–50 640, May 2022.

[39] R. Agravat and M. S. Raval, "3D Semantic Segmentation of Brain Tumor for Overall Survival Prediction," Oct. 2020.

[40] G. Garg and R. Garg, "Brain Tumor Detection and Classification based on Hybrid Ensemble Classifier," Jan. 2021.

[41] S. Bhuvaji, A. Kadam, P. Bhumkar, S. Dedge, and S. Kanchan, "Brain Tumor Classification (MRI)," [Online]. Available: https://www.kaggle.com/dsv/1183165, 2020.

[42] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," Computer Vision and Pattern Recognition, 2015.

[43] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 10.1109/CVPR.2016.30, 2015.

[44] F. Chollet, "Xception: Deep Learning With Depthwise Separable Convolutions," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[45] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2015.

[46] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," In. Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016.

[47] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. "Inception-v4, inception-ResNet and the impact of residual connections on learning," In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17), San Francisco, California, USA, 2017

[48] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, San Diego, California, 2016.

[49] Z. Jia and D. Chen, "Brain tumor identification and classification of MRI images using deep learning techniques." IEEE Access, August 2020.

[50] Subhankar Roy, Willi Menapace, Sebastiaan Oei, Ben Luijten, Enrico Fini, Cristiano Saltori, et al., "Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound," in IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2676-2687, doi: 10.1109/TMI.2020.2994459, Aug. 2020.

[51] Qing Lyu, Hongming Shan, Yibin Xie, Alan C. Kwan, Yuka Otaki, Keiichiro Kuronuma, Debiao Li, and Ge Wang, "Cine Cardiac MRI Motion Artifact Reduction Using a Recurrent Neural Network," in IEEE Transactions on Medical Imaging, vol. 40, no. 8, pp. 2170-2181, Aug. 2021.

[52] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021), "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Los Alamitos, CA, USA, 2021.

[53] Hassani A, Walton S, Shah N, Abuduweili A, Li J, Shi H. "Escaping the Big Data Paradigm with Compact Transformers", Apr. 2021.

[54] M. -H. Guo, Z. -N. Liu, T. -J. Mu and S. -M. Hu, "Beyond Self-Attention: External Attention Using Two Linear Layers for Visual Tasks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.