Article - Engineering, Technology and Techniques

MRI Brain Tumor Classification Using a Hybrid VGG16-NADE Model

Saran Raj Sowrirajan[1*]
https://orcid.org/0000-0002-5954-5924

Surendiran Balasubramanian[2]
https://orcid.org/0000-0001-5435-0880

Raja Soosaimarian Peter Raj[3]
https://orcid.org/0000-0002-7216-2207

[1]National Institute of Technology Puducherry, Department of Computer Science and Engineering, Karaikal, Puducherry, India;[2]National Institute of Technology Puducherry, Department of, Computer Science and Engineering, Karaikal, Puducherry, India; [3]Vellore Institute of Technology, School of Computer Science and Engineering, Vellore,Tamilnadu, India.

*Correspondence: sarandilip.er@gmail.com (S.R.S.).

HIGHLIGHTS

- Proposed a hybrid VGG16_NADE model

- Additional evaluation metrics such as the macro-F1 and weighted F1 values

- Offering enhanced performance to handle problems such as identifying leaf disease and damaged paddy seeds.

Abstract: A brain tumour is determined to be abnormal cell development on the brain walls and inside the skull. A malignant variation is a dangerous form of cancer with an increased mortality rate. Analyzing Magnetic Resonance Imaging (MRI) through deep learning models is the most prevalent and accurate method of early cancer detection. A novel hybrid model is proposed with the VGG16 convolution neural network (CNN) and Neural Autoregressive Distribution Estimation (NADE). The experiment was conducted on 3064 MRI brain tumour images grouped into three categories. The T1 weighted contrast-enhanced MRI images were classified using the hybrid VGG16-NADE model and compared with other methods. The results prove that the proposed hybrid VGG16-NADEmodel outperforms the rest in terms of classification accuracy, specificity, sensitivity and F1 score. The prediction accuracy of the proposed hybrid VGG16-NADE is 96.01%, precision 95.72%, recall 95.64%, F-measure 95.68%, Receiver operating characteristic (ROC) 0.91, error rate 0.075, and the Matthews correlation coefficient (MCC) 0.3564. The numerical outcomes are comparatively higher than those from other approaches and it is evaluated with existing approaches like the hybrid CNN and NADE, CNN, CNN- kernel Extreme Learning Machines (KELM), deep CNN-data augmentation, and CNN- Genetic Algorithm (GA). Other metrics like the p-value, MCC, error rate and ROC are also evaluated. The experimental outcomes show that the hybrid VGG16-NADE classifier model outperforms other approaches.

INTRODUCTION

A brain tumour is a mass that develops either on the brain or skull walls and is of two types, benign and malignant. Such a development of masses known as tumours occurs at random in the brain and affects the body. Early diagnosis and detection of brain tumours play a critical role in their cure. To make this happen, radiologists use MRI images of the brain to detect abnormal cell growth and identify the stages of cancer. Advancements in computer-aided diagnosis have automated MRI image analysis, while machine learning (ML) and deep learning (DL) methods have helped detect brain tumours accurately. The basic steps involved in brain tumour detection are preprocessing, feature extraction and classification [1].

Different imaging techniques such as the Positron emission tomography (PET), Computed tomography CT and MRIare used to screen for brain tumours. APET scan injects a radioactive tracer to identify the quantum of chemical activity in the diseased portion. A CT scan forms cross-sectional images of the brain by revolving an X-ray tube around a patient's body and releasing a narrow X-ray beam towards the body. An MRI scan applies a powerful magnetic field directed towards the patient's body to arrange protons in the body prior to transmitting radiofrequency signals through the body. The protons release energy and coordinate with the magnetic feld when there is no power. The response of multiple brain tissues is recorded by the energy released from the image. The two types of MRI are functional (fMRI) and structural (sMRI). In the former, brain activity is examined, depending on the variations in the blood flow. In the latter, the anatomy and pathology of the brain are recorded. The current research utilizes the sMRI to handle pathology in the brain. The sMRI extracts tissue responses, which may result in varied biological information in the image. The different sMRI are discussed below:

The Diffusion Weighted Image (DWI) MRI measures water molecule diffusion using MR imaging and frequently envisages hyperintensities.

– The Fluid-attenuated inversion recovery (FLAIR) MRI is an MRI pulse sequence that reduces fluid while improving oedema.

– The T1w MRI measures the longitudinal relaxation time (time taken by the protons to return to equilibrium), where the variations in tissue are captured using a standard MRI pulse sequence.

– The T1Gd MRI helps obtain theT1 sequence following the introduction of the contrast-enhancing agent, gadolinium, into the body. It reduces T1 time, and brightens the appearance of blood vessels and tumours.

– The T2w MRI extracts differences in the tissue transverse relation time (T2) using a standard MRI pulse sequence.

More recently, research on brain tumour detection has been carried out through deep learning methods, producing much more successful rates with large datasets. Computer-aided diagnostic systems have overcome the flaws of the previous systems with the introduction of deep learning methods. Computer-aided diagnostic systems have overcome the flaws of the previous systems with the introduction of deep learning methods. Today, convolution neural networks have taken the medical world by storm, especially in the detection of brain tumours. The convolution layers in the network extract features from the given MRI images using a number of differently-sized filters [1]. Pretrained CNNs, Alexnet, GooglNet, Lenet, and ResNet are used for a slew of classification problems [2]. A recent study was undertaken with a hybrid model, using different CNNs for the classification problem, to maximize performance. AlexNet with the Long short-term memory (LSTM) model and ResNet with the LSTM model produced 71% accuracy. VGGNet with the LSTMoffered 84% accuracy in the previous study with the MRI brain tumour dataset, while performance increased by 95%through transfer learning in the Resnet50 model. The VGG16 architecture outperforms other kinds in brain tumour segmentation by using the Mask R-CNN with transfer learning[3]. Clearly, the VGG16 works better when the model is utilized as a hybrid. The CNN seems to have a higher computational cost, owing to its multiple convolution layers. A hybrid model was designed to classify the tumour accurately by extracting supplementary information and patterns from the MR image, with limited CNN layers to reduce the computational cost. The CNN model learns additional information from thegiven MR images by estimating the probability of pixel information so as to classify the tumour accurately.

Data distribution in machine learning is a challenge. There are definite correlations of each image pixel to its neighbour pixel that are tough to estimate when there is no prior knowledge about the images. The Auto-Regressive (AR) model is a popular estimator method that determines correlations among image data. The resultant image from the AR model is free from noise and redundancies in the image[4]. A dataset with a few training samples, tumours in different shapes, and unbalanced data of all classes is recommended to determine the distribution of brain tumours from MRI brain images. The Neural Autoregressive Distribution Estimation(NADE) is an outstanding pixel density estimator refined from the Restricted Boltzmann

Machines[5]. The model can find the density information of real-valued data, binary data, and other CNN architecture like the VGGNet[6]. The NADE model is obtained from axiomatic solutions that successfully identify valid joint distribution. Deep neural networks, which are ideally suited to learning intricate information from data, support accurate classification and recommender systems[7]. This article proposed different learning methods to construct a hybrid architecture by applying the properties of VGG16 and NADE architecture. Feature extraction using the VGG16 and density estimation using NADE provide additional information for accurate learning. The experiment carried out with the T1-weightedcontrast-enhanced MR brain tumour dataset produced 97.31% classification accuracy. The major research contributions of the proposed approach are listed below:

Implementing a hybrid VGG16_NADE using the VGG16 to extract detailed features from the MRI dataset, and highlighting the features by smoothening the tumour border and removing redundant content from the MRI.

Measuring performance by considering additional evaluation metrics such as the macro-F1 and weighted F1 values.

Offering enhanced performance to handle problems such as identifying leaf disease and damaged paddy seeds.

Tiwari and coauthors [8] discussed, numerous image processing approaches have been applied to improve MR image features for the identification and classification of tumor tissue. Image segmentation with the fuzzy C-means and k-means provide useful features from image data. When it comes to evaluating and comparing images, image segmentation is crucial. Tissue classification, identification, and estimating the tumour area are only a few of its uses in brain imaging. Gumaste and coauthors [9] discussed statistical properties like homogeneity, mean and absolute values of MR images were evaluated using the Support Vector Machine (SVM) classifier. Gumaei and coauthors[10] Employing algorithms such as the principal component analysis (PCA), speeded up robust features (SURF) descriptors, and others, much research includes a feature extraction phase to extract features with the most critical information. Sharif and coauthors[11]discussed brain tissue classification was undertaken using extreme learning, after a hybrid feature extraction method based on the covariance matrix selects a combination of features using the particle swarm optimization technique. Tandel and coauthors[12] discussed ML techniques like the Random Forest (RF), Decision Tree (DT), Naive Bayes(NB), and SVM were analyzed to assess the performance of the proposed technique. Kaldera and coauthors[13] demonstrated that, the appearance of malignant cells is a significant challenge in brain tumour identification and classification. For each malignant tumour form, the size, shape, location, and intensity of the malignant tissue varies from image to image. Convolutional neural networks (CNN) have, in recent times, gained popularity for feature extraction, particularly in medical images, and for video analyses as well. The ability of the CNN to predict fundamental patterns from training images is its most important feature. Bhanothu and coauthors[14] introduced an automated brain tumour identification and classification system using a faster R-CNN method. The VGG16 has been chosen as the framework for the faster R-CNN to obtain a feature vector that highlights the tumour region and classifies tumour types in MRI brain tumour images. Precision is considered the evaluation metric in the proposed work, and the average precision achieved for all classes of tumours is 77.60%.

Alkassar and coauthors[15] proposed a scheme that segments the tumour region in the MR brain image by employing the VGG16 and full convolution network properties. Tumour regions are segmented automatically by applying the proposed model, which achieved 97% accuracy with the BRAST2015 dataset. Naser and coauthors[16] proposed the pre-trained CNN VGG16 and U-net are adopted to simultaneously segment the tumour region and rank low-grade glioma. The proposed method obtained accuracy of 84% for segmentation, 92% for tumour detection, and 89% for tumour grading. Kang and coauthors[17] proposed an ensemble method for brain tumour classification using MRI brain tumour data. The preprocessing techniques utilized on input images are followed by different pre-trained CNN models used with various machine learning classifiers. The CNN extracts in-depth features, relying on the convolution layers, following which the data is forwarded to the classification layer. This study proved that a pre-trained model with large convolution layers can extract in-depth features from images and is suitable for extensive datasets with three or four classes.

Badža and coauthors[18] designed new CNN with max-pooling and dropout layers, improving network performance during the training process. The network was tested using the T1contrast-enhanced MRI dataset. The performance metrics are evaluated using both original and augmented data with ten-fold cross-validation. The architecture was evaluated for average precision, recall, accuracy and F1-score and obtained 95.40% accuracy. Rai and coauthors[19] proposed a less complex and low-layer U-net CNN for an MRI brain tumour dataset, consisting of tumour and non-tumour classes. The architecture is evaluated through metrics like specificity, precision, F-score, recall and accuracy for performance accuracy and obtained a score of 98%. Sajjad and coauthors [20] proposed a novel CNN model for brain tumour classification. MRI data is

given as input to the CNN model after preprocessing, resizing, cropping and augmenting the input image. Before training the model, the tumour in the MRI brain image is segmented and its spatial information augmented. The performance accuracy is measured, and it received a score of 94.58%. Montúfar and coauthors[21]discussed Restricted Boltzmann Machines and variational auto-encoders are latent variable generation models that allow assumptions of independence to decrease the number of factors and parameters. The visible variables in RBMs are considered independent, given the hidden variables, permitting the use of block Gibbs sampling. Models that do not include any latent variables are autoregressive (AR) models, which aim to compute a typical distribution over the input image. The models use default factorization by the chain rule but do not include any latent variables, while NADE offers data density information.

Janowczyk and coauthors [22] disucussed semantic segmentation, particularly of organs, figures prominently in image analysis. CNN, and especially DNN architecture, has been popular since 1990 with the two-layer LeNetarchitecture. However, the development of Alex Net with its five convolutional layers and accessibility to fast GPUs as well as related computational facilities over fifteen years radically changed the CNN landscape. Multiple layers such as convolution, non-linearity, pooling, regularisation, optimization and normalization, loss functions, and network parameter initializations are utilized in developing CNNs. Xu and coauthors [23]used a double-pathway CNN with 2D multi-resolution input patches to execute convolution operations and integrate path way outputs. The DeepMedicCNN also includes double paths with 3D multi-resolution input patches and integrated residual connections. The U-net comprises an encoder-decoder structure with skip connections, and the serial ensemble technique is used in the anisotropic framework. The entire tumour is partitioned in the initial network, and the tumour core is partitioned in the second, depending on the previous outcome. The developing tumour is partitioned in the final layer, using the last result received. A 2D Fully Connected Neural Networks (FCNN) technique with the CRF is utilized, and the FCNNs and residual connections are used for segmentation. The first FCNN partitions the entire tumour, while the second FCNN segregates the interior sections of the tumour. Sun and coauthors[24] introduced FCNN-based encoder-decoder architecture for partitioning distinct tumour sub-areas. The three different FCNN designs deployed demonstrate that those with multi-resolution features outperform the ones with single-resolution structures. Further, the Dilated Residual Network that is developed is provided with similar patches for training.

Tran and coauthors[25]employed DeepMedic, FCN, and U-net to create seven networks with three different preprocessing algorithms. The final label is generated from the results of each layer, after the preprocessing is completed. Dense and dilated modules are created in the encoder-decoder cascaded design to broaden the research. Dilated convolution layers are used instead of pooling layers. A combination of 10 encoder-decoder-based frameworks is constructed, including an auto-encoder stream, to recreate and regularise the actual image. Zhang and coauthors[26]advanced a new method that combines the U-net and DenseNet. The network is expanded to form three networks, based on three different image perspectives. Zhao and coauthors[27] modelled a two-network cascaded route. The first is a rough partitioning network that splits the WT and the second is an exemplary partitioning network that divides the network's sub-areas. The two-network model utilized a four-level deep 3D U-net topology.

Wang and coauthors[28]developed a combination of six 3D U-net models with varying input sizes, feature maps and encoding/decoding blocks. After the z-score is standardized, features are input to the network. Lekadir and coauthors[29]developed an FCN and acquired results for three axes, with the final partitioning results obtained using the highest vote. Ten features are created from the outcomes, and the mean PCA and SD-PCA are utilized to apply the RF. A group consisting of the DFKZNet, U-net and Cascaded Anisotropic (CA)-CNN, along with the RF and its 14 radionics features, is chosen from multiple Laplacian, Gaussian and wavelet-decomposed images. A 2D U-net framework is designed for tumour partition. Finally, OS prediction is carried out using characteristics like the age, volume, and shape of the entire tumour. With the recent neuroimaging techniques MRI brain scan process is performed and these techniques include functional MRI, traditional structural MRI, diffusion tensor imaging, and diffusion weighted imaging. The structural MRI process intends to differentiate abnormal and healthy brain tissues using molecule content which is generally employed in standard imaging techniques [30]. This process assists in visualizing healthy brain tissues and maps radiation-induced micro haemorrhage, calcification, tumour vascularity, and mapping gross brain anatomy. The structural technique includes FLAIR, T1-w and T2-w and contrast-enhanced T1-w. On the other hand, the functional MRI is adopted to acquire the neural activity inside the brain via the oxygenated ratio to deoxygenate the blood level with the neighbourhood vasculature while evaluating the motor or cognitive task [31]. The fMRI is utilized to differentiate tumour grades and localize eloquent cortex. DWI acquires the random motion of water molecules over the brain and it is utilized to

characterize tumour via the prediction of hypoxia and cellularity, WM tract integrity, per-tumoraledema and differentiate posterior fossa tumours [32]. However, diffusion tensor imaging (DTI) is utilized to examine 3D diffusion direction and it is also known as diffusion tensor of water molecules. DTI assists in determining local tumour effects on white matter tract integrity with tract destruction, tumour infiltration, vasogenicedema existence and tract displacement[33-35].

MATERIAL AND METHODS

This research applies data augmentation and learning processes over an MRI brain tumour data set available online. Here, the samples are trained using the hybrid NADE and VGG16 model and compared with various existing approaches like the standard CNN, CNN with the KELM, deep CNN with data augmentation, and CNN with the GA. The dataset is split into two segments for testing and training. The testing data is considered for the final evaluation of the hybrid NADE and VGG16, and the training data is evaluated for model learning. The proposed model is composed of diverse phases, and an over view of the model is shown in Figure [5].

The dataset

This work considers the brain T1-weighted CE-MR brain dataset, collected from Nanfang Hospital from 2005 to 2010, shown in Fig 1. It is composed of 3064 slices from 233 different patients. The dimensions of the slices are 512*512 pixels of 0.49 mm * 0.49 mm. The dataset has three kinds of tumours, namely, pituitary tumours, gliomas and meningiomas, as seen in Table 1. Three expert radiologists examine the patients' pathology reports to ascertain their pathology type and image labels. Here, the images are handled independently, and the radiologists can reach a consensus in regard to the tumour label over every image. In this work, two tumour images of the same category are determined based on their relevance, i.e. similarity and dissimilarity.

Table 1. Image dataset summary

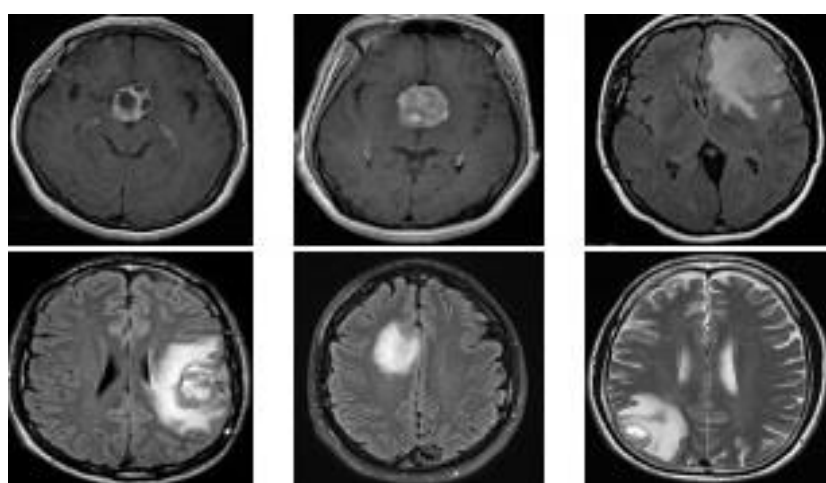| Category	Total patients	No. of slices	View	No. of slices
Meningiomas	82	708	Transverse	209
			Sagittal	231
			Coronal	268
Gliomas	89	1426	Transverse	494
			Sagittal	495
			Coronal	437
Pituitary tumours	62	930	Transverse	291
			Sagittal	320
			Coronal	319

Figure 1. Sample MR images fromthe T1-weightedcontrast-enhancedMR braintumour dataset[32].
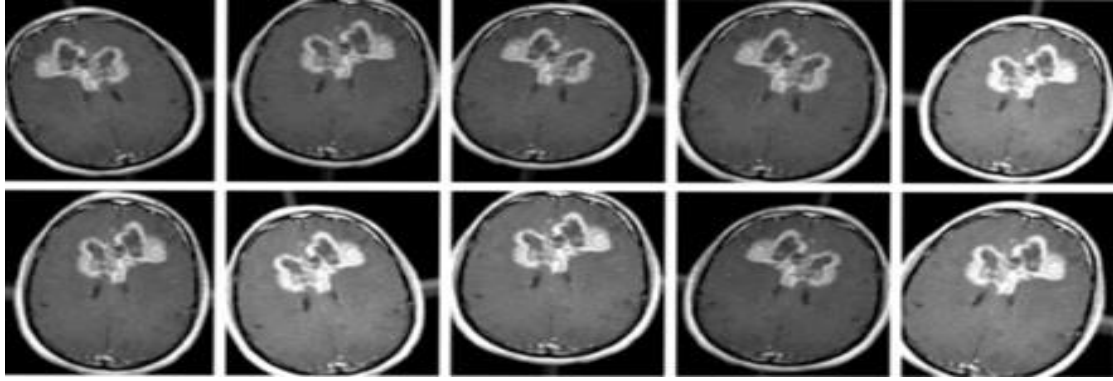
Figure 2. Augmented MRI images

Data augmentation

Data augmentation is used to increase the dataset volume artificially, based on the volume of existing data. The training process with DL concepts requires a considerable quantum of data with fine-tuned parameters. Given that the samples from the proposed dataset are few, this research uses data augmentation for the dataset training process with minor changes in terms of brightness, rotation and flipping. Consequently, the size of the training data increases, with these minor variations being considered distinct images and facilitating superior model learning with the unseen data. Figure 2 depicts the augmented image.

Neural Autoregressive Distribution Estimator (NADE)

The NADE process is initiated with a D-dimensional distribution-based $p(x)$ observation, which is factored into the product of 1D-distribution with an order of integer permutation, as expressed in Equation (1):

$$p(x) = \prod_{d=1}^{D} p(x_{o_d}|x_{o<d}) \tag{1}$$

Here, $o < d$ comprises the initial $d-1$ dimensional while ordering o, and $x_{o<d}$ is related to the sub-vector for all the dimensions. Therefore, the 'auto-regressive' generative model that is defined is based on data simplification with the specification of D conditional parameterization $p(x_{o_d}|x_{o<d})$. Every condition is modelled with a feed-forward neural network (FFNN) in the proposed NADE. Especially, every $p(x_{o_d}|x_{o<d})$ condition is parameterized, as in Equation (2) and Equation (3):

$$p(x_{o_d} = 1|x_{o<d}) = sigm(v_{o_d,}h_d + b_{o_d}) \tag{2}$$

$$h_d = sigm(W_{.,o<d}x_{o<d} + c) \tag{3}$$

Here, $sigm(a) = 1/(1 + e^{-a})$ denotes the logistic sigmoid, H the number of hidden units, and $V \in \mathbb{R}^{D*H}, b \in \mathbb{R}^D, W \in \mathbb{R}^{H*D}, c \in \mathbb{R}^H$ the NADE parameters. The bias c and hidden layer matrix W are provided for every hidden layer, HD (with the same size). Since the parameter-sharing scheme specifies that the NADE model must possess $O(HD)$ parameters, $O(HD^2)$ is essential when the NN is separate. The number of parameters is limited to diminish the risk of overfitting. A significant advantage of this model is that all the D hidden layers, h_d, are evaluated in the $O(HD)$ time of the $O(HD^2)$. The pre-activation of the dth hidden layers is represented as $a_d = W_{.,o<d}x_{o<d} + c$ and the complexity is obtained using the recurrence given in Equation (4) and Equation (5) below:

$$h_1 = sigm(a_1); \quad where a_1 = c \tag{4}$$

$$h_d = sigm(a_d); \quad where a_d = W_{.,o<d}x_{o<d} + c = W_{.,o_{d-1}}x_{o_{d-1}+a_{d-1}} for all d$$
$$\in \{2,...,D\} \tag{5}$$

From Equation (5), the vector a_{d-1}formed is computed with $O(H)$. However, the computation of Equation (2) with $h$is $O(H)$.Therefore, the calculation of $p(x)$ from the D conditional distribution is $O(HD)$ for NADE[36]. The complexity is measured with various standard FFNN models. Figure 3 depicts the Neural Autoregressive Distribution Estimator (NADE) applied on the T1-weighted contrast-enhanced MRI images.
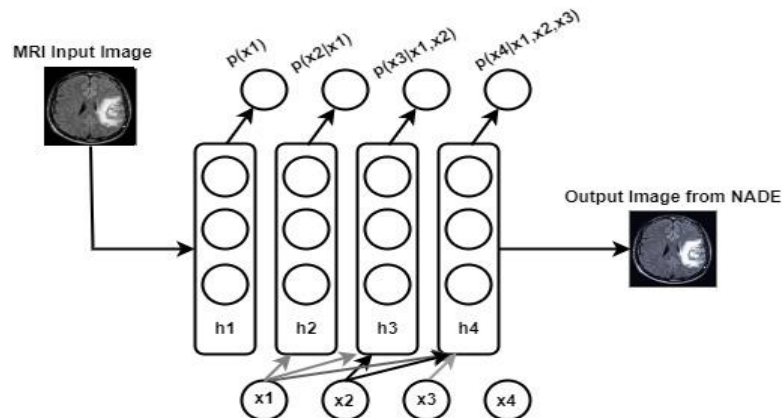

Figure 3. Neural Autoregressive Distribution Estimator applied on the T1-weighted contrast-enhanced MRI images.

NADE is trained using the maximal likelihood or its equivalent by reducing the average negative log-likelihood, expressed in Equation (6):

$$\frac{1}{N}\sum_{n=1}^{N} -\log p\big(x^{(n)}\big) = \frac{1}{N}\sum_{n=1}^{N}\sum_{d=1}^{D} -\log p\left(x_{o_d}^{(n)}\Big|x_{o<d}^{(n)}\right) \tag{6}$$

The model uses the stochastic gradient descent (mini-batch). Given the probability $p(x)$and cost $O(HD)$,the gradients of the negative log-probability of the training samples are evaluated in $O(HD)$.

VGG16 Architecture

VGG16 architectureis the most outstanding vision network architecture around. The unique feature ofthe VGG16 is its focuson using 3x3 filter convolution layers rather than building massive hyperparameters. The architecture has been designed with the same padding and 2x2 filter maxpoolinglayers. This arrangement of convolution layers followed by the max poolinglayers is the same throughout the entire architecture. Initially, the network consists of 64 single filters sized 3x3 to detect edges, corners, and lines in images. The max pooling layer of kernel size 2x2 is added, followed by the convolution layer, so as to get a complete summary of values from the image array. This arrangement is continued throughout the entire architecture with 128,256 and 512 filters sized 3x3 at each convolution layer[37]. ReLu is the linear activation function placed at every convolution layer to activate the neurons in the network. Two fully connected layers with the ReLU activation function are added to the architecture, followed by the softmax function to classify the brain tumour. The VGG16 is a most extensive network with 138-plus million parameters. The addition of more convolution layers helps the VGG16 learn hidden features from the image. The network input is a image dimension (224, 224,3). The initial layers (two) pose 64 channels with 3*3 filter size with same padding. After the stride of max pooling layer (2,2), two layers have convolution layers with 256 filter size and (3,3) filter size. It is followed by the maximal pooling layer of stride (2,2) which is same alike of prior layer. Then, there exist two sets of 3 convolution layer with filter size (3,3) and 256 filter. Then, there are two sets of convolution layer and max pool layer. Each poses 512 filters of (3,3) size with same padding. The image is then provided to the stacked convolution layers (two). In this max pooling and convolution layers, 3*3 filters are used indeed of 11*11. In some layers, it uses 1*1 pixel which is utilized for manipulating the number of input channels. There is a padding 1-pixel (similar padding) performed after every convolution layer to avoid the spatial image feature. After the stacking max-pooling and convolution, there are (7,7,512) feature maps are attained. The output is flattened with feature vectors. Then, there are three FL layers, and the first layer considers the input form the last feature vector and provides a (1,4096) output vector, and the second layer works alike of first layer. However, the third layer provides an output in various channels which is passed to the softmax layer to normalize the classification vector. The hidden layer adopts ReLU as activation function which is computationally effectual. However, it outcomes in faster learning process and reduces the likelihood of vanishing gradient issues. VGG-16 is a superior architecture and provides better prediction accuracy.Fig 4 depicts the VGG16 architecture.
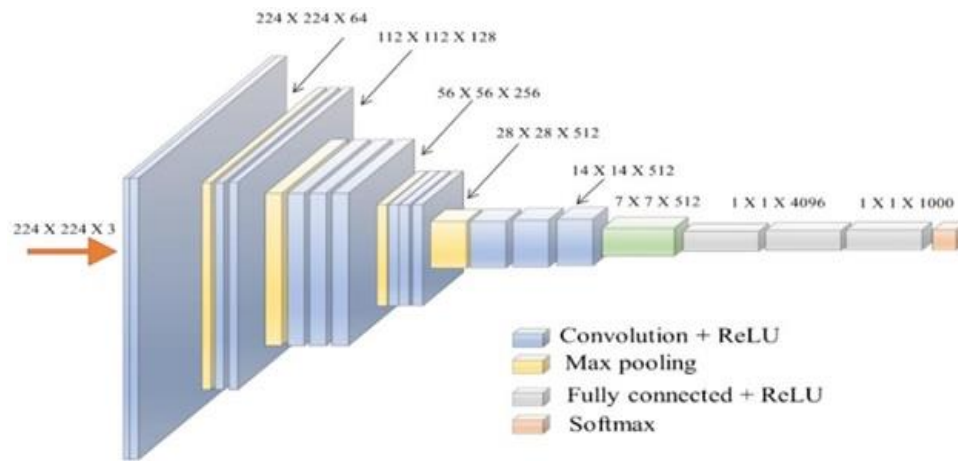
Figure 4. VGG16 architecture

Convolutional module

Here, the convolutional model is considered the 4th pooling layer, and its scaling invariant helps capture image clues Features appropriate for MR imagesare extracted from the pooling layer, while those acquired from successive lower or higher layers are not, as the images are either specific or generic. Therefore, the outcomes from the pooling layer are connected to this module

Fully connected layer(FCL)

The 1D image is transformed with the concatenated features from the convolutional/maxpooling layers. FCLs, which are used explicitly for this purpose, comprise three layers known as dense, drop out and flatten. Here, the drop out is fixed as 0.5 and the dense layer as 256.

Softmax layer

The softmax layer is utilized to classifyfeatures from the FC layer, which is also known as the dense layer, where the unit number is based on the number of categories. It outputs the multi-nominal probability distribution score that relies on the classification process. The output distribution is expressed in Equation (7):

$$P(a = c|b) = \frac{e^{b_k}}{\sum_j e^{b_j}}$$
(7)

Here, b and c specify the probabilities retrieved from the softmax layer and classes of the MR dataset used. An architectural description of the proposed model is provided in Table 2.

Table 2. VGG16 Layer description

Input	Layers	Feature mapping	Size	Kernel size	Stride	Activation
1	2*conv	1	224*224*3	-	-	-
2	Max pool	64	224*224*64	3*3	1	ReLU
3	2*conv	64	224*224*64	3*3	2	ReLU
4	Max pool	128	112*112*64	3*3	1	ReLU
5	2*conv	128	112*112*128	3*3	2	ReLU
6	Max pool	256	56*56*128	3*3	1	ReLU
7	3*conv	256	28*28*256	3*3	2	ReLU
8	Max pool	512	28*28*512	3*3	1	ReLU
9	3*conv	512	14*14*512	3*3	2	ReLU
10	Max pool	512	14*14*512	3*3	1	ReLU
11	FC	512	7*7*512	3*3	2	ReLU
12	FC	-	25088	-	-	ReLU
13	FC	-	4096	-	-	ReLU
14	FC	-	4096	-	-	ReLU
Output	FC	-	1000	-	-	Softmax

Proposed Hybrid VGG16-NADE Model

The hybrid architecture consists of the NADE, followed by the VGG16 and the output layer. Input MRI brain images are first taken to train the NADE. The forward propagation technique used in NADE architecture learns features, estimates probability density, and finds joint distribution. A new model image is generated by removing redundancies and smoothing the tumour border in the MRI brain image. The NADE helps the CNN learn the tumour region, which has different shapes and locations. A newly generated image from the NADE is given as input to the VGG16, which is a robust network that extracts detailed features from the input data (See Fig 5). The VGG16is a pre-trained convolution neural network consisting of 64,128,256 and 512 3x3 filters followed by the maxpooling layers. The FCL at the end of the convolution layer mapsthe features into a one-dimensionalfeature vector so as to categorize the input data intotumour classes. The Cross_Entropy_Loss function is applied to evaluate the loss, and is represented by Eq. (8):

$$H(y, \hat{y}) = \sum_i y_i log \frac{1}{\hat{y}_i} = -\sum_i y_i log \hat{y} \tag{8}$$
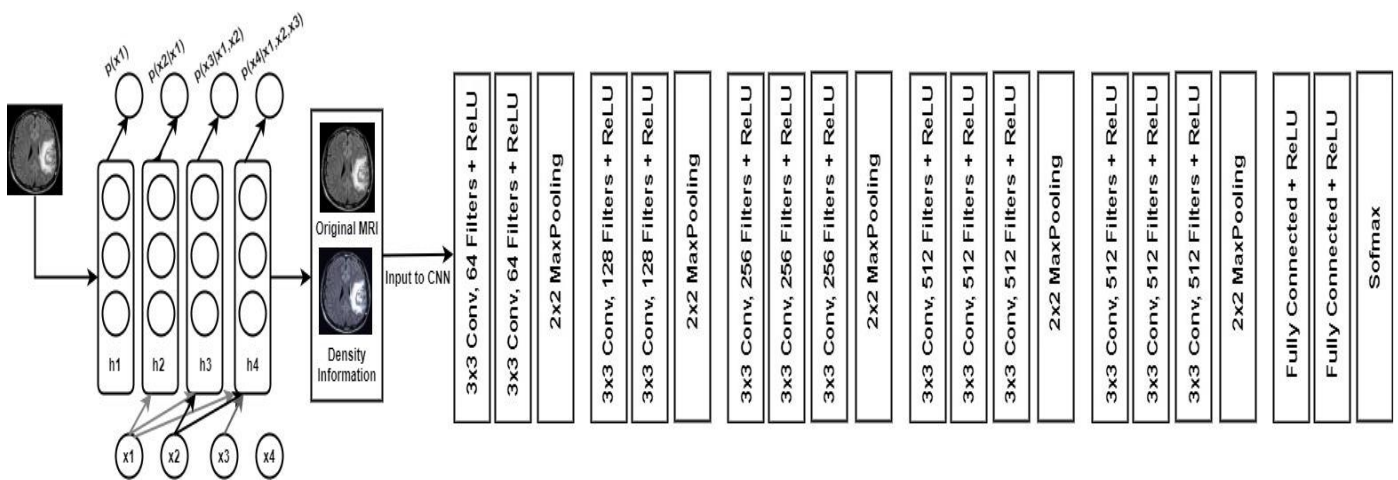

Figure 5. Proposed hybrid VGG16 and Neural Autoregressive Distribution Estimator (NADE) architecture

To reduce the loss involved during training the model, parameters such as learning rates and weights are adjusted. The ADAM optimizer that is used for the learning process combines the gradient descent and RMSprop optimizer. The Adam optimizer is implemented using the weighted mean of the past gradient and the weighted mean of its square to update the network's weight and bias.

The autoregressive density estimator model suffers from high-dimensional data, and input data contains plenty of input features. The CNN is robust architecture that extracts features with high-dimensional data, while the NADEdoes so by finding the joint probability distribution over the images. The convolutional NADE extractsfeatures after removing redundant portions in the image. The architecture of the proposed hybrid VGG16-NADE model consists of the NADE, followed by the VGG16, which is a pre-trained CNN with an output layer. The model usesthe relevant region of the tumour images for a comprehensive analysis. Here, the NADE is hybridized with the VGG16, and the performance requires dealing with hundreds of hidden units. The model is trained with contrastive divergence and computes the distribution estimators. The partition function is not precisely evaluatedbut approximated using the sampling process. The hybrid model measures the mean with the unbounded weight using the empirical samples. The accurate test log-based likelihood averages are less than the errors and bias values. The NADE-based hidden layers possess a learning rate of 0.0005 with a constant decrease rate of -88.8, which is the best for the average likelihood and superior to the NADE training with fewer steps. The overall functionality of the hybridized NADE-VGG16 model is enhanced with training using the stochastic gradient descent to produce 96.01% prediction accuracy.

The model shows improvedgeneralization with no cost in terms of tumour prediction. It is observed that the enhanced performance of the model depends on better optimization with the stochastic gradient descent. The log-likelihood of the NADE is -84, which is near the test log-likelihood. Further, the NADE can take advantage of the non-linear optimization method. On the contrary, training the VGG16 requires that the gradients approximate certain sampling values. The VGG16 is severely restricted to a simple optimization approach. Samples are generated from the NADE with VGG16training and sequential sampling, according to the $p(v_i|V_i)$estimated by the hybridized model. The samples are provided with the probability used to acquire them. The final images are clear, and the exacted samples are scrutinized under the NADE. The

hybridized model is thus used to evaluate the distribution of high-dimensional samples. The proposed hybridized NADE and the VGG16 model outperform various existing approaches in terms of tumour prediction.

EXPERIMENTAL RESULTS AND DISCUSSION

This section discusses the numerical outcomes of the proposed hybridized model. A simulation is carried out with the MATLAB 2020a environment where metrics like precision, accuracy, recall and F-measure are computed. Input images of different dimensions are resized to 224x224 pixels before being fed into the hybrid architecture. To begin with, the NADE model eliminates redundancies in the image and smoothens the tumour border to assist the VGG16 network extract essential features from the MRI brain images. The VGG16 contains 13 convolution layers of 64,128,256 and 512 channels with 3x3 filters for each layer. The max pooling layer has 2x2 filters to reduce the image size by 2, and the softmax function is used in the output layer to classify brain abnormalities. The optimization technique used in the network is the loss function (cross-entropy).During the training process, 10-fold cross-validation is used for a performance evaluation of the proposed architecture, given its extensive data size. Performance measured with the original MRI brain image dataset is 93.57 in the VGG16, and 96.01% in the VGG16 after combining the original data with data density information. A confusion matrix is provided to calculate the metrics of each model, including the F1-score, precision, and sensitivity, and performance s evaluated through the confusionmatrix for classificationproblems.

Accuracy

Accuracy refers to the number of input patterns identified correctly over a set of data instances, and is measured by Equation (9):

$$Accuracy = \frac{TrueNegative + TruePositive}{TrueNegative + FalsePositive + TruePositive + FalseNegative} \quad (9)$$

F1-score

It is the harmonic mean of recall and precision, evaluated by Equation (10):

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (10)$$

Precision

Precision is a positive predictive value expected to be 1 for a good classifier. The false positive rate must be zero for precision, which is evaluated by Equation (11):

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (11)$$

Recall

Recall is a positive predictive value, expected to be 1 for a good classifier. The falsenegative rate must be zerofor recall, which is evaluated by Equation (12):

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (12)$$

Weighted F1

The number of samples from each class is used for the weighted-F1 score, which is calculated using Equation (13):

$$Weighted\ F1 = \frac{1}{No.ofSamples} \sum_{Classes} No.ofsamplesfromclassXF1 \\ - Scoreoftheclass \quad (13)$$

MacroF1

The classifier's overall F1 score is calculated by merging all the classes of F1 scores into a single value. The macro F1-score is calculated by Equation (14):

$$MacroF1 - score = \frac{1}{No.ofClasses} \sum_{i=0}^{Classes} F1 - score_i \qquad (14)$$

Macro-precision

Precision is calculated for each class(PrC1, PrC2 and PrC3) and macro-precision is calculated using Equation (15):

$$Macro - precision = \frac{PrC1 + PrC2 + PrC3}{Totalno.ofClasses} \qquad (15)$$

Specificity

Specificity identifies instances appropriately. It is estimated by calculating the fraction of true negatives in healthy samples, and is expressed mathematically in Equation (16) as:

$$Specificity = \frac{TN}{TN + TP} \qquad (16)$$

Macro-sensitivity

Sensitivity is calculated for each class(SnC1, SnC2 and SnC3), and macro-sensitivity is calculated by taking the mean of the overall sensitivity. It is expressed thus in Equation (17):

$$Macro - sensitivity = \frac{SnC1 + SnC2 + SnC3}{Totalno.ofClasses} \qquad (17)$$

The performance of the proposed approach is superior to that of other approaches which consist of limited layers in the CNN. The dataset, whichcomprises sizes 224x224 for the proposed architecture, extracts detailed features from the MRI. Such a process shows that increasing the number of layers, as well as that of filters, with different dimensions in the CNN maximizes system performance. Future directions may include the use of pre-trained CNN architecture with additional layers, like the DenseNet and VGG19,forimproved performance.

Confusion matrix

Also known as the error matrix,the confusion matrix is utilized to see if the classification result matches the original ground cover. It is the foundation for several additional assessment criteria, and is expressed in Equation (18).

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1c} \\ \vdots & \ddots & \vdots \\ x_{c1} & \cdots & x_{cc} \end{bmatrix} \qquad (18)$$

Wheren_{ij} = number of subcategories and x_{ij} = number of categories. The element on the diagonal indicates several successfully split data points and $(i, j = 1,2, ..., n)$ the number of samples that splits into the n-th category (absolute number). Here, 'n' is the total number of sample elements.

$$n = \sum_{i=1}^{c} \sum_{j=1}^{c} x_{ij} \qquad (19)$$

The Mathews Correlation Coefficient (MCC) is evaluated withthe true and predicted classes as binary values. The critical propertyrelies on a valuebetween -1 and +1.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FN)}} \qquad (20)$$

The Mean Square Error (MSE) estimates the average of the squares of error difference between the actual and the estimated value.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y}_i)^2 \qquad (21)$$

Here, y_i is the original value, $\bar{y}_i$ the predicted value, and N the data points.

Table 3. Confusion matrix for the hybrid VGG16-NADEmodel

Predict	Meningioma	Glioma	Pituitary
Meningioma	652	49	7
Glioma	53	1368	5
Pituitary	4	4	922

Table 4. Metrics of the hybrid VGG16-NADEmodel

Macro-Sensitivity	Specificity	Macro-Precision	Macro-F1	Weighted-F1
95.64	97.90	95.72	95.68	96.01

Table 5. A comparison of the F1 score for the VGG16, Three_Layer CNN and hybrid VGG16-NADE models

	VGG 16	Three_Layer CNN	Proposed VGG16-NADE
Meningioma	89.01	88.07	92.02
Glioma	94.70	93.65	96.10
Pituitary	97.61	96.15	98.92

Table 6. A comparison of various performance metrics

Methods	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)	ROC	Error rate	MCC
Hybrid CNN and NADE	95	94.49	94.21	94.56	0.86	0.089	0.4656
CNN	96.13	95.61	95.62	95.46	0.855	0.074	0.4534
CNN and KELM	93.6	83.4	76.5	72	0.84	1.56	0.4317
Deep CNN with data augmentation	94.58	81.2	75	77.9	0.80	1.63	0.4658
CNN and Genetic Algorithm	94.2	91.97	93.62	94.24	0.78	0.077	0.4205
U-net Architecture with CNN Model	95.89	95.55	95.52	95.53	0.89	0.077	0.4016
ED-GAN	92.35	91.66	92.33	91.66	0.76	1.35	0.4642
GAN + ConvNet	95.6	95.29	94.91	95.10	0.90	0.076	0.4156
Proposed Hybrid VGG16-NADE	96.01	95.72	95.64	95.68	0.91	0.075	0.3564

Table 3 depict confusion matrix and Table 4 depict the evaluation of metrics like macro-sensitivity, macro-precision, specificity, macroF1, and weightedF1. The macro-sensitivity of the VGG16_NADE is 95.64%, specificity 97.90%, macro-precision 95.72%, macroF1 95.68%, and weightedF1 96.01%. The F1-score of the proposed VGG16-NADE is compared with that of the standard VGG16 and the three_layer CNN model in Table 5. In the case of meningioma, the F1-score of the VGG16-NADE is 92.02%, which is 3.01% and 3.95% higher, respectively, than that of the VGG16 and the three-layer CNN. In the case of glioma, the F1-score of the VGG16-NADE is 96.10%, which is 1.4% and 2.45% higher, respectively, than that of the VGG16 and the three_layer CNN. In the case of the pituitary, the F1-score of the VGG16-NADE is 98.92%, which is 1.31% and 2.77% higher, respectively, than that of the VGG16 and the three-layer CNN.

Table 6 compares various performancemetrics like prediction accuracy, recall, precision, F-measure, ROC, error rate and MCC. The prediction accuracy of the hybrid VGG16-NADE is 96.01%, which is 1.01%, 2.41%, 1.43%, 1.81% and 0.51% higher, respectively, than that of the hybrid CNN and NADE, CNN and KELM, deep CNN with data augmentation, CNN-GA, and U-Net with CNN. The precision of the hybrid VGG16-NADE is 95.72% which is 1.23%, 0.11%, 12.32%, 14.52%, 3.75%, and 0.17 higher, respectively, than that of thehybrid CNN and NADE, CNN, CNN and KELM, deep CNN with data augmentation, CNN-GA,

and U-Net with CNN. The recall of the hybrid VGG16-NADE is 95.64% which is 1.43%, 0.02%, 19.14%, 20.64%, 2.02%, and 0.12 higher, respectively, than the hybrid CNN and NADE, CNN, CNN and KELM, deep CNN with data augmentation CNN-GA and, U-Net with CNN. The F-measure of the hybrid VGG16-NADE is 95.68% which is 1.12%, 0.22%, 23.68%, 17.78%, 1.44%, and 0.15 higher, respectively, than the hybrid CNN and NADE, CNN, CNN and KELM, deep CNN with data augmentation, CNN-GA and, U-Net with CNN. The ROC of the proposed VGG16-NADE is 0.91, which is comparatively higher than that of other approaches at 0.05, 0.0055, 0.07, 0.11, 0.13, and 0.02 respectively. The error rate of the proposed model is 0.075, which is less than that of other approaches. Generally, the MCC value should range from -1 to 1; however, the proposed model gives a better MCC value of 0.3564 while the others offer values of 0.4656, 0.4534, 0.4317, 0.4658, 0.4205, and 0.416, respectively. NADE shows huge advantages by evaluating the probability of every pixel where resultant image attained is smoothening border and eliminates anomalies over the brain tumours. However, this cannot be effectually attained by both the GAN and other attention mechanism. While discussing about GAN with a sample, the image can be produced based on some general patterns and it is not so appropriate for image smoothening and outlier or anomaly prediction. Another drawback associated with GAN is its training iterations. The time consumption is higher while handling the large size images and the resulting output gives lesser image resolution.

Table 7 depicts the training accuracy and training loss of the proposed classifier model with ten epochs. The model shows 96.87% training accuracy with ten epochs and 8.65% training loss. Similarly, the validation accuracy for ten epochs is 98% and training loss is 2.35%, which is substantially better for the proposed model (See Table 7 and Table 8).

Table 7. Measuring training accuracy and loss

Epochs	1	2	3	4	5	6	7	8	9	10
Training accuracy	73.71	82.051	81.651	88.661	91.471	92.981	93.891	95.134	95.592	96.875
Training loss	51.61	39.084	31.812	25.677	20.977	17.667	15.221	12.555	11.661	8.654

Table 8. Measuring validation accuracy and loss

Epochs	1	2	3	4	5	6	7	8	9	10
Validation accuracy	85.50	90.51	96.51	93.51	96.51	96.51	98.5	100	99.1	98.1
Validation Loss	32.23	24.26	8.63	15.05	8.74	6.87	4.36	1.83	2.36	2.36

P-value computation

The Wilcoxon ranksum method is applied to evaluate the proposed approach. According to this method, the performance score is normalized to [0, 1] and the null hypothesis is tested, based on the p-value.

Table 9. Statistical computation

Significance $\gamma = 0.01$	
Comparison	$p - Value$
$VGG - 16\ and NADE$	0.0060

If the p-value is less than the threshold level, it is superior to other algorithms; and if it is equal to the threshold level, the validity of the model is negligible, making it equivalent to the different approaches discussed above.

If the p-value is rather more significant, the algorithm is statistically not superior. From the statistical analysis (SeeTable 9), it is clear that the proposedmodel shows a better p-value of 0.0060 in regard tothe conventional classifier for a significant value of 0.01. The analysis proves that the model works effectually to measure the performance of the classifier, and outperforms existing approaches efficiently with superior prediction accuracy

Table 10. A performance comparison with various CNN methods

Dataset	ML algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score	Time (min)	AUROC
Kaggle for brain tumour classification	Hybrid CNN and NADE	92.54	0.74	0.70	0.14	0.10	92.86
	CNN	87.87	0.77	0.68	0.13	0.10	92.36
	CNN and KELM	88.98	0.71	0.71	0.14	0.78	93.86
	Deep CNN with data augmentation	86.76	0.74	0.72	0.14	6.80	91.40
	CNN and Genetic Algorithm	87.87	0.85	0.89	0.15	1.100	92.32
	Standard CNN	93.14	0.90	0.92	0.16	0.06	95.89
	Proposed VGG16-NADE	95.87	0.91	0.94	0.20	0.04	96.78
OpenfMRI dataset	Hybrid CNN and NADE	90.13	0.73	0.68	0.12	0.05	91.50
	CNN	91.12	0.75	0.67	0.12	0.08	91.14
	CNN and KELM	90.39	0.76	0.69	0.12	0.78	90.14
	Deep CNN with data augmentation	91.80	0.74	0.67	0.12	6.06	89.38
	CNN and Genetic Algorithm	90.13	0.70	0.71	0.12	1.62	91.97
	Standard CNN	93.15	0.91	0.92	0.20	0.04	92.56
	Proposed VGG16-NADE	95.56	0.92	0.93	0.24	0.04	93.58
T1-weighted-Contrast-Enhanced-MR Images brain-tumour Dataset	Hybrid CNN and NADE	95	94.49	94.21	94.56	1.25	86
	CNN	96.13	95.61	95.62	95.46	2.658	85.5
	CNN and KELM	93.6	83.4	76.5	72	4.56	84
	Deep CNN with data augmentation	94.58	81.2	75	77.9	3.56	80
	CNN and Genetic Algorithm	94.2	91.97	93.62	94.24	4.57	78
	Standard CNN	90.1	89.56	85	90	2.89	89
	Proposed VGG16-NADE	96.01	95.72	95.64	95.68	0.04	91

Table 10 depicts an overall comparison of the proposedand existing approaches. In the case of the Kaggle dataset for brain tumour classification, the performance of the proposed model is evaluated with prevailing approaches like the hybrid CNN and NADE, CNN, CNN and KELM, deep CNN with data augmentation, CNN+GA, and standard CNN. The accuracy of the proposedmodel is 95.87%, which is 3.33%, 8%, 6.89%, 9.11%, 8%, and 2.73% higher, respectively,than that of other approaches. The precision of the proposed model is 91%, which is 17%, 14%, 20%, 17%, 6% and 1% higher,respectively,than that of other approaches. The recall of the proposedmodel is 94%, which is 24%, 26%, 23%, 22%, 5% and 2% higher,respectively, than that of other approaches. The F1-score of the proposedmodel is 20, which is comparatively higher than that of other approaches. The execution time of the proposed model is 0.04 min, which is substantially superior with 96.78% AUROC. In the open MRI dataset, the accuracy of the proposedmodel is 95.56%, which is 5.43%, 4.44%, 5.17%, 3.67%, 5.43% and 2.41% higher, respectively,than that of other approaches. The precision of theproposed model is 92%, which is 19%, 17%, 16%, 18%, 22% and 1% higher, respectively,than that of other approaches. The recall of theproposed model is 93%, which is 25%, 26%, 24%, 26%, 22% and 1% higher, respectively, than that of other approaches. The F1-score of the proposed model is 24, which is comparatively higher than that of other approaches. The execution time is 0.04 min, superior to that of other models with 93.58% AUROC. In the case of the T1-weighted contrast-enhanced MR images dataset, the model accuracy is 96.01% which is 1.01%, 2.41%, 1.43%, 1.81% and 5.91% higher, respectively,than that of other approaches. The precision of the proposed model is 95.72% which is 1.23%, 0.11%, 12.32%, 14.52%, 3.75% and 6.16% higher, respectively,than that of other approaches. The recall of the proposed anticipated model is 95.64% which is 1.43%, 0.02%, 19.14%, 20.64%, 2.02% and 10.64% higher,respectively, than that of other approaches. The F1-score of the proposed model is 95.68, which is comparatively higher than that of other approaches. The execution time is 0.04 min, which is superior to that of other models with 93.58% AUROC. Based on this analysis, it is demonstrated that the model works most effectually with the datasetprovided in regard to all metrics.

Table 11. A performance evaluationbased on data partition

Metrics	Data Splitting Range (%)	VGG16-NADE	VGG16	Three_Layer CNN
Accuracy	30-70	72.18	68.27	68.18
	35-65	74.90	72.92	71.95
	40-60	79.46	74.80	74.25
	45-55	82.50	78.97	76.78
	50-50	86.55	81.08	80.01
	55-45	90.12	84.26	83.20
	60-50	93.15	87.97	85.18
	65-55	95.14	92.45	88.98
	70-30	96.01	93.57	92.22
Precision	30-70	69.87	69.87	67.25
	35-65	71.25	70.19	69.82
	40-60	74.52	72.68	70.21
	45-55	76.89	76.16	75.27
	50-50	79.81	81.02	81.97
	55-45	83.89	85.56	84.51
	60-50	87.19	89.87	87.97
	65-55	93.68	91.71	90.02
	70-30	95.72	94.38	91.17
Recall	30-70	68.97	67.46	69.46
	35-65	72.50	73.15	69.91
	40-60	75.25	76.45	72.34
	45-55	76.15	78.25	75.96
	50-50	78.56	81.86	80.99
	55-45	84.56	85.87	85.79
	60-50	88.98	87.89	88.78
	65-55	94.19	91.92	91.91
	70-30	95.64	93.57	90.28
F-measure	30-70	70.21	71.91	72.48
	35-65	73.68	74.65	75.31
	40-60	74.49	77.79	77.45
	45-55	77.21	80.67	81.26
	50-50	79.67	84.58	83.13
	55-45	84.72	87.91	86.03
	60-50	89.91	90.74	88.19
	65-55	93.46	92.68	92.57
	70-30	95.68	94.38	90.46

Table 11 depicts a performance evaluation based on data partitioning, that is, 30% testing and 70% training, 35% testing and 65% training, 40% testing and 60% training, 45% testing and 55% training, 50% testing and 50 training, 55% testing and 45% training, 60% testing and 50% training, 65% testing and 55% training, 70% testing and 30% testing samples, respectively. The proposed model gives higher prediction accuracy with 96.01%, which is substantially superior to that of the VGG16 and the three-layer CNN. Here, the data partitioning rate is 70% for training and 30% for testing. Similarly, the precision, recall and F1-measure of the VGG16-NADE 95.72%, 95.64% and 95.68, which is higher than that of the VGG16 and the three-layer CNN. The model shows better results not only for 70:30, but also for 30:70, 35:65, 40:60, 45:55, 50:50, 55:45, 60:50, 65:55, and 70:30 samples. The prediction accuracy of the VGG16-NADE is 96.01%, which is 2.44% and 3.79% higher than that of the VGG16 and the three-layer CNN for the 70% training and 30% testing samples. The precision of the VGG16-NADE is 95.72%, which is 1.34% and 4.55% higher than that of the VGG16 and the three-layer CNN for the 70% training and 30% testing samples. The recall of the VGG16-NADE is 95.64%, which is 2.07% and 5.36% higher than that of the VGG16 and three-layer CNN for 70% training and 30% testing samples. The F-measure of the VGG16-NADE is 95.68%, which is 1.3% and 5.22% higher than that of the VGG16 and three_layer_CNN for 70% training and 30% testing samples. Based on the analysis, it is evident that the model works most effectually to predict brain tumor classification.

Table 12. A Comparison of proposed method by varying Network parameters and without Augmentation techniques

Performance Metrics	VGG16-NADE with Augmentation (proposed)	VGG16-NADE without Augmentation	Proposed method with change of kernel size 1 in 7th, 10th and 13th layer in network
True Positive	2942	2829	2937
True Negative	6006	5893	6001
False Positive	122	235	127
False Negative	122	235	127
Accuracy	96.01	92.33	95.86
Precision	95.72	92.03	95.33
Recall	95.64	91.40	95.66
F-Score	95.68	91.68	95.66

Table 12 depicts a comparison of proposed method by varying network parameters and without Augmentation techniques.

CNN efficiently performs on large-scale data. The network trained with few number samples leads to poor performance. Applying the data generation method to the proposed model improves the performance. In proposed VGG16-NADE, layers parameter values are modified in the 7, 10, and 13th layers of the network. Choosing a small kernel value will see detailed depth information in the image. In this experiment, the total number of trainable parameters was reduced to 1.2 million. So, the time duration of training the model is reduced but does not outperforms the proposed model.

CONCLUSION

A hybrid VGG16_NADE model has been proposed for brain tumour classification. The VGG16 pre-trained CNN consists of 16 layers that extract detailed features from MRI images. The Adam optimizer and softmax function used in the architecture undertake learning and classification. A neural autoregressive density estimator (NADE) removes redundant brain images and smoothens the tumour border. The NADE gives the VGG16additional density information from the MRI images during the learning process. The proposed method is tested with a T1-weighted contrast-enhancedMR image braintumor dataset, and is trained usingthe Adam optimizer and 10-fold cross-validation. The performance is evaluated with macro-sensitivity, macro-precision, specificity, macroF1, weightedF1, F1-score, and 96.01% accuracy is achieved. The model is compared for accuracy with five current methods, and the study shows that the hybrid method is the most preferred for medical image applications.The prediction accuracy of the hybrid VGG16-NADE is 96.01%, precision 95.72%, recall 95.64%, F-measure 95.68%,ROC 0.91, error rate 0.075, and the MCC 0.3564, which is comparatively higher than that of other approaches.The major research constraint is the ability of VGG-16 to train the dataset, i.e. slow to train and this drawback is due to the size of the input which occupies huge disk space and bandwidth makes it ineffectual.

Conflicts of Interest: The authors declare no conflict of interest

REFERENCES

1. Díaz-Pernas FJ, Martínez-Zarzuela M, Antón-Rodríguez M, González-Ortega D. A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network. Healthcare. 2021 Feb;9(2):153.
2. Mehrotra R, Ansari MA, Agrawal R, Anand RS. A transfer learning approach for AI-based classification of brain tumors. Mach. Learn. With Appl. 2020 Dec;15(2):100003.
3. Shahzadi I, Tang TB, Meriadeau F, Quyyum A.CNN-LSTM: cascaded framework for brain tumour classification. 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES); 2018 Dec 3; Sarawak, Malaysia:IEEE; 2018.p. 633-7.
4. Oliva J, Dubey A, Zaheer M, Poczos B, Salakhutdinov R, Xing E, Schneider J. Transformation autoregressive networks. International Conference on Machine Learning; 2018 Jul 3;Stockholm, Sweden: PMLR;2018.p. 3898-907.
5. Khoshaman A, Vinci W, Denis B, Andriyash E, Sadeghi H, Amin MH.Quantum variational autoencoder. Quantum Sci. Technol. 2018; 4(1):1–13.
6. Uria B, Murray I, Larochelle H.RNADE: The real-valued neural autoregressive density-estimator, Adv. Neural Inf. Process. Syst. 2013:26(1):1–12.
7. Zhang S, Yao L, Sun A, Tay Y. Deep learning based recommender system: A survey and new perspectives. ACM Comput. Surv. 2019; 52(1):1–35.

8.   Tiwari A, Srivastava S, Pant M. Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019. Pattern Recognit. Lett. 2020;131(1): 244–60.
9.   Gumaste PP, Bairagi VK. A hybrid method for brain tumor detection using advanced textural feature extraction. Biomed. Pharmacol. J. 2020; 13(1):145–57.
10. Gumaei A, Hassan MM, Hassan MR, Alelaiwi A, Fortino G. A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification. IEEE Access. 2019;7(1):36266–73.
11. Sharif M, Amin J, Raza M, Yasmin M, Satapathy SC. An integrated design of particle swarm optimization (PSO) with fusion of features for detection of brain tumor. Pattern Recognit. Lett. 2020;129(1):150–7.
12. Tandel GS, Biswas M, Kakde OG, Tiwari A, Suri HS, Turk M, et al. A review on a deep learning perspective in brain cancer classification.Cancers. 2019;11(1):111.
13. Kaldera HN, Gunasekara SR, Dissanayake MB. Brain tumor classification and segmentation using faster R-CNN. 2019 Advances in Science and Engineering Technology International Conferences (ASET); 2019 Mar 26; Dubai, United Arab Emirates. IEEE; 2019.p. 1–6.
14. Bhanothu Y, Kamalakannan A, Rajamanickam G. Detection and classification of brain tumor in MRI images using deep convolutional network. 2020 6th international conference on advanced computing and communication systems (ICACCS); 2020 Mar 6; Coimbatore, India. IEEE; 2020. p. 248–252.
15. Alkassar S, Abdullah MA, Jebur BA. Automatic brain tumour segmentation using fully convolution network and transfer learning. 2019 2nd International Conference on Electrical, Communication, Computer, Power and Control Engineering (ICECCPCE); 2019 Feb 13; Mosul, Iraq. IEEE; 2019. p. 188-192.
16. Naser MA, Deen MJ. Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images. Comput. Biol. Med. 2020 Jun 1;121(1):103758.
17. Kang J, Ullah Z, Gwak J. Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers.Sensors. 2021 Mar 22;21(6):2222.
18. Badža MM, Barjaktarović MČ. Classification of brain tumors from MRI images using a convolutional neural network. Applied Sciences. 2020 Mar 15;10(6):1999.
19. Rai HM, Chatterjee K. Detection of brain abnormality by a novel Lu-Net deep neural CNN model from MR images. Machine Learning with Applications. 2020 Dec 15;2(1):100004.
20. Sajjad M, Khan S, Muhammad K, Wu W, Ullah A, Baik SW. Multi-grade brain tumor classification using deep CNN with extensive data augmentation.J Comput.Sci. 2019 Jan 1;30(1):174-82.
21. Montúfar G. Restricted boltzmann machines: Introduction and review. Information Geometry and Its Applications IV. Springer, Cham. 2016 Jun 12; 252(1):75-115
22. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J Pathol Inform.2016 Jan 1;7(1):29.
23. Xu Y, Mo T, Feng Q, Zhong P, Lai M, Eric I, Chang C. Deep learning of feature representation with multiple instance learning for medical image analysis. In2014 IEEE international conference on acoustics, speech and signal processing (ICASSP); 2014 May 4; Florence, Italy. IEEE; 2014. p. 1626–1630.
24. Sun W, Tseng TL, Zhang J, Qian W. Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. Comput Med Imaging Gr. 2017 Apr 1; 57(1):4–9.
25. Tran PV. A fully convolutional neural network for cardiac segmentation in short-axis MRI. arxiv: 1604.00494. abs/1604.00494. 2016 Apr 2;1(1):1-21.
26. Zhang H, Li L, Qiao K, Wang L, Yan B, Li L, et al. Image prediction for limited-angle tomography via deep learning with convolutional neural network. arXiv preprint arXiv:1607.08707. 2016 Jul 29;1(1):1-21.
27. Zhao J, Zhang M, Zhou Z, Chu J, Cao F. Automatic detection and classification of leukocytes using convolutional neural networks. Med Biol Eng Comput. 2017 Aug;55(8):1287-301.
28. Wang S, Yao J, Xu Z, Huang J. Subtype cell detection with an accelerated deep convolution neural network.International Conference on Medical Image Computing and Computer-Assisted Intervention; 2016 Oct 17; Athens, Greece.Springer, Cham. 2016. p. 640–648.
29. Lekadir K, Galimzianova A, Betriu À, del Mar Vila M, Igual L, Rubin DL, et al. A convolutional neural network for automatic characterization of plaque composition in carotid ultrasound. IEEE J Biomed Health Inform. 2016 Nov 22;21(1):48-55.
30. Zhou T, Canu S, Ruan S. Fusion based on attention mechanism and context constraint for multi-modal brain tumor segmentation. Computerized Medical Imaging and Graphics. 2020 Dec 1;86(1):101811.
31. Naser MA, Deen MJ. Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images. Computers in biology and medicine. 2020 Jun 1;121(1):103758.
32. Zhang J, Zeng J, Qin P, Zhao L. Brain tumor segmentation of multi-modality MR images via triple intersecting U-Nets. Neurocomputing. 2021 Jan 15;421(1):195-209.
33. Quon JL, Bala W, Chen LC, Wright J, Kim LH, Han M, et al. Deep learning for pediatric posterior fossa tumor detection and classification: a multi-institutional study. Am. J. Neuroradiol. 2020 Sep 1;41(9):1718-25.
34. Varuna Shree N, Kumar TN. Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network. Brain informatics. 2018 Mar;5(1):23-30.
35. Ismael MR, Abdel-Qader I. Brain tumor classification via statistical features and back-propagation neural network. 2018 IEEE international conference on electro/information technology (EIT); Rochester, MI, USA. IEEE;2018:252–7.

36. Larochelle H, Murray I. The neural autoregressive distribution estimator. The fourteenth international conference on artificial intelligence and statistics; Florida, USA. PMRL; 2011. p. 29-37.
37. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014 Sep 4;1(1):1-14.