Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

Brain Tumor Detection and Classification from MRI Images using Fuzzy Thresholding, a Convolutional Neural Network and Grad-CAM Explanation

Pooja A
ME Communication System
Dept. of Electronics and Communication Engineering.
PET Engineering College,Valliyur, Tirunelveli

C S Sree Thayanandeswari, M.E (Phd)
Assistant professor
Dept. of Electronics and Communication Engineering.
PET Engineering College,Valliyur, Tirunelveli

K. Bhagya Shree, M.E
Assistant professor
Dept. of Electronics and Communication Engineering.
PET Engineering College,Valliyur, Tirunelveli

Abstract: Brain tumor is a serious disease and finding it early helps a lot in planning the treatment. Reading MRI scans by hand takes time and it depends on how experienced the doctor is. In this paper we present a system that detects and classifies brain tumors from MRI images using fuzzy thresholding, a convolutional neural network (CNN) and Grad-CAM explanation. First the images are converted to grayscale, resized, filtered with a median filter and enhanced with CLAHE. Then fuzzy c-means thresholding picks out the suspected tumor region and the tumor area is estimated from the mask. A small CNN with four convolutional blocks and about 0.39 million learnable parameters classifies each image into glioma, meningioma, pituitary tumor or no tumor, and Grad-CAM draws a heatmap on the tumor region so that the doctor can see why the network gave that answer. The system was built in MATLAB and tried on a dataset of 5589 MRI images. The CNN reached an overall accuracy of 71.93% with an average precision of 75.99%, recall of 71.98% and F1-score of 70.26%. The no-tumor class was classified best (F1 88.37%) while meningioma was often mixed up with pituitary (recall 32.29%). When the three tumor classes are merged into one, the same predictions separate tumor from no tumor with 94.31% accuracy. The fuzzy segmentation and the heatmap worked well on the sample shown, but the four-class classifier needs more training before it can be trusted, and we discuss the reasons, the limits of the evaluation and the next steps.

Keywords: Brain Tumor, MRI, Fuzzy Thresholding, Fuzzy C-Means, Convolutional Neural Network, Grad-CAM, Explainable AI, Medical Image Classification, Tumor Segmentation

I. INTRODUCTION

A. Background

A brain tumor is a group of cells that grow in an abnormal way inside or near the brain. Some tumors grow slowly and

some grow fast, and what they do to the patient depends on the type, the size and the place where they sit [1]. The skull is a closed space, so even a tumor that is not cancerous can press on the nearby tissue and cause headaches, seizures, vision problems or changes in memory and behaviour. Because of this, finding the tumor early and knowing its type is very important for choosing the treatment, and the type decides if the doctor will go for surgery, radiation, drugs or just careful follow-up [2].

MRI is the scan most often used for this job. It uses a strong magnetic field and radio waves, it shows the soft tissue of the brain in good detail, and it does not use ionizing radiation. A single patient can have many slices in different planes and with different sequences, so a radiologist has a lot of images to go through. The reading is slow and it depends on the experience of the person. Two doctors can also disagree on a difficult slice, mainly when the tumor is small or when its edge is blurred.

B. Motivation

Computer aided methods can help here. Earlier works used hand made features, for example texture features from wavelets [3], hybrid intelligent techniques [4] or local binary patterns [5], together with a classifier such as SVM [6]. Later, deep learning models were used which learn the features by themselves [7], [8]. These networks give good accuracy, but many of them only say a class name. A doctor has no way to check whether the network looked at the tumor or at something else in the image, like the skull or an artefact. In medicine this is a real problem, because a prediction that cannot be checked is hard to trust [9], [10].

Another point is that most of the strong models are big. They have many millions of parameters, they need a lot of data and a good GPU, and they are not easy to put in a small device. A small network which is cheap to train and run, and which still shows its reasoning, is therefore interesting, even if its accuracy is a bit lower to start with. This paper follows that direction.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

C. Problem statement

The problem we look at is the following. Given a brain MRI slice, decide whether a tumor is present and, if yes, which type it is (glioma, meningioma or pituitary), find where the tumor is, estimate its size, and show the doctor the reason for the decision. Each of these parts has been solved alone in the literature, but they are seldom put together in one light system.

D. Objectives

The objectives of this work are: (1) to improve the quality of the MRI image with simple pre-processing; (2) to extract the suspected tumor region with fuzzy thresholding, which handles the blurred tumor border; (3) to classify the image into four classes with a small CNN; (4) to explain the decision with a Grad-CAM heatmap that stays on the tumor; and (5) to measure the performance honestly with the confusion matrix and class-wise measures.

E. Contributions

The main points of this work are listed below.

- **Combined pipeline:** pre-processing, fuzzy c-means tumor extraction, a four-class CNN and Grad-CAM are connected in one MATLAB system that gives the class, the tumor area and a visual explanation.
- **Tumor mask and size:** a mask generation procedure (brain mask, membership threshold, region selection and morphological clean-up) is described step by step, and the tumor area is estimated in mm^2 from the mask.
- **Explanation limited to the tumor:** the Grad-CAM map is multiplied with the fuzzy tumor mask, so the heatmap stays on the tumor region and is easier for a doctor to check.
- **Detailed error analysis:** besides the usual class-wise measures, we study the row-normalised confusion matrix, the kappa value and the tumor versus no-tumor view, and we show that most of the errors come from one pair of classes.
- **Honest reporting:** the weak points of the classifier and of the evaluation are stated openly, so that the next version can fix them.

F. Organization of the paper

The rest of the paper is arranged like this. Section II gives the medical and technical background. Section III reviews the related work. Section IV explains the proposed method in detail. Section V gives the results and the discussion. Section VI lists the limitations, Section VII talks about ethical and clinical points, and Section VIII concludes the paper and describes the future work.

II. BACKGROUND

A. Types of brain tumor

Brain tumors are mainly grouped by the type of cell they start from and by how they grow [2]. Four common types are shown in Fig. 1 and summarised in Table I. The dataset used in this paper has three of them (glioma, meningioma and pituitary tumor) and one more class for images without a tumor.

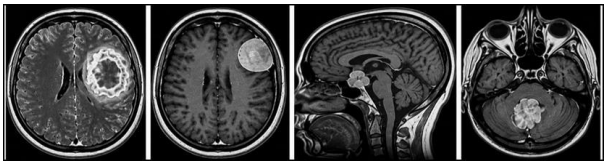


Fig. 1. Examples of brain tumor types (glioma, meningioma, pituitary tumor, medulloblastoma).

TABLE I. COMMON BRAIN TUMOR TYPES

Type	Where it starts	Remarks
Glioma	Glial cells of the brain	Can grow slowly or fast, depends on grade
Meningioma	Meninges (layers around brain and spinal cord)	Mostly slow growing, problem when it presses nearby tissue
Pituitary tumor	Pituitary gland at the base of the brain	Can change hormone levels and affect vision near optic nerves
Medulloblastoma	Usually the cerebellum	Malignant, mostly seen in children
No tumor	–	Images without the target abnormality, lets the model learn normal appearance

The reasons for a brain tumor are different for each type. Genetic changes, inherited conditions and radiation exposure may play a role, but many times the exact reason is not known [1]. The symptoms also depend on the type and place of the tumor. They include headaches which do not go away or become worse, seizures, vomiting, vision problems, trouble with balance, changes in memory or thinking, and weakness or numbness in parts of the body. Since these signs are not special to tumors, a scan is needed to be sure.

B. Brain MRI

Brain MRI gives clear information about grey matter, white matter, the ventricles and abnormal regions, and it is used to find tumors, bleeding, swelling and infection. Different sequences, such as T1-weighted, T2-weighted, FLAIR and DWI, show the tissue in different ways (Fig. 2). The same brain can be imaged in the axial, sagittal and coronal planes (Fig. 3), and each plane helps in a different way: the axial plane is good for the size and shape of a tumor, the sagittal plane for the position and extent, and the coronal plane for the relation to nearby structures (Table II).

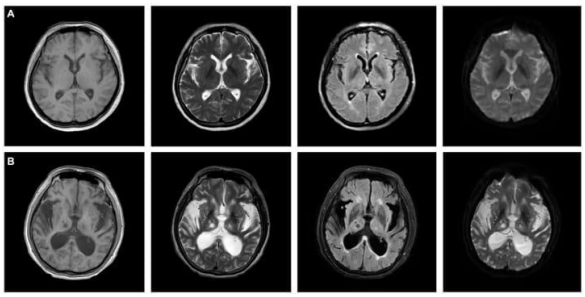


Fig. 2. Brain MRI in different sequences.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

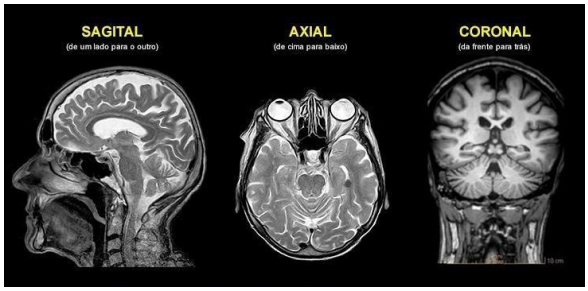


Fig. 3. MRI views: sagittal, axial and coronal.

TABLE II. MRI VIEWS AND THEIR USE

View	Direction	Useful for
Axial	Horizontal cross-section	Size and shape of the tumor, comparing it with nearby tissue
Sagittal	Side view, front to back	Position and extent of the tumor, internal structures
Coronal	Front view, left to right	Relation of the tumor to nearby structures, spread of tissue

C. Challenges in MRI image analysis

Automatic analysis of brain MRI is not easy, and the main difficulties are listed in Table III together with the step of our system which tries to deal with each. The intensity of the images changes from scanner to scanner. Tumors have many sizes and shapes. The border of the tumor is often not clear, and sometimes the tumor has nearly the same intensity as the normal tissue. Noise and artefacts make all this worse, and a model trained on one dataset may not work on another one.

TABLE III. CHALLENGES AND HOW THIS WORK HANDLES THEM

Challenge	Step used in this work
Noise and artefacts	3×3 median filter
Low or uneven contrast	CLAHE
Unclear tumor border	Fuzzy memberships instead of a hard threshold
Different tumor size and shape	Convolutional features, rotation/shift/flip augmentation
Black-box decision	Grad-CAM heatmap restricted to tumor mask
Differences between datasets	Not solved here, listed as future work

D. Artificial intelligence in medical diagnosis

Artificial intelligence can read a large number of medical images in a short time and find patterns linked to a disease. In image analysis it supports pre-processing, segmentation, detection and classification. Machine learning and deep learning models learn the patterns from the data, so the manual work of designing features is reduced, and the results are more consistent than those of a tired human reader. But there are limits too. The model needs enough good and correctly labelled data. A model trained on one scanner or hospital may fail on another. Deep models need computing power, and many of them are hard to interpret. Most important, an AI output should not be taken as a final diagnosis; it must be checked by a clinician after proper testing and validation [10].

E. Fuzzy thresholding and explainability in short

Two ideas from outside deep learning are used in this paper. The first one is fuzzy thresholding. In normal thresholding a pixel is either tumor or not tumor. In an MRI the tumor border is blurred, so a pixel at the border is partly tumor and partly normal. With fuzzy clustering, each pixel gets a membership value between 0 and 1 for every cluster, so these partial cases are kept instead of being cut [11], [12]. The second idea is explainable AI (XAI). XAI methods try to show why a model gave an answer. For images, a common way is to colour the regions which had the strongest effect on the output; Grad-CAM does this using the gradients flowing into the last convolutional layer [13].

F. Why several evaluation measures are needed

Accuracy is the share of images classified correctly, and it is easy to understand, but it can hide problems. If one class is much bigger than the others, a model can get a high accuracy by favouring that class. Even with balanced classes, as here, accuracy does not tell which classes fail. Precision answers the question “when the system says this class, how often is it right?”, and recall answers “of the images of this class, how many did the system find?”. In medicine the two have different meanings: a low recall means missed patients, and a low precision means false alarms. The F1-score joins both in one number, and the confusion matrix shows exactly which classes are mixed up. For this reason we report all of them, and not only the accuracy.

III. LITERATURE REVIEW

A lot of work has been done on brain tumor detection from MRI. We group it here by the main idea used, and at the end we point out what is still missing.

A. Classical image processing and machine learning

The early systems first extracted features by hand and then used a classifier. Sasikala and Kumaravel selected a wavelet based texture feature set for tumor classification [3], and El-Dahshan et al. mixed several intelligent techniques for classifying MRI images [4]. Kaplan et al. tried modified local binary pattern features [5], and Rao and Karunakara used a kernel based SVM [6]. These methods are light and their steps are easy to explain, but they depend on the features chosen by the designer. When the tumors differ much in appearance, a fixed feature set may miss the important pattern, and the accuracy often drops on new data.

B. CNN and transfer learning for classification

With deep learning the feature design is done by the network. Badža and Barjaktarović built a CNN for classifying three tumor types from MRI [7], and Hammad et al. proposed a light end-to-end model for tumor detection [8]. Pashaei et al. combined a CNN with extreme learning machines [14], and Rasheed et al. showed that image enhancement before the CNN helps [15]. Mohsen et al. used deep neural networks on wavelet features [16]. Transfer learning with big pre-trained networks also became common: Soumik and Hossain used an Inception network [17], Noreen et al. joined the features of Inception-V3 and DenseNet201 before a Softmax classifier [18], Majib et al. used a VGG16 based network with augmentation [19], Shah et al. fine-tuned EfficientNet [20], and Younis et al. fine-tuned ResNet50 for four classes [21]. Rizwan et al. proposed a Gaussian CNN for tumor and glioma grade classification [22]. These models show that transfer

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

learning is strong, but they are heavy; for example VGG16 has a large number of parameters, and two backbones used together double the memory need. They also mostly give a label only and say nothing about where the tumor is.

C. Hybrid, attention and transformer models

More recent works join several ideas. Chaki et al. used fuzzy logic for image enhancement followed by an Inception-ResNet based network with reinforcement learning, and reported class accuracies between about 94% and 100% for meningioma, glioma, pituitary and no-tumor images [23]. Preetha et al. combined a three-branch CNN with EfficientNetB2 and reported about 99.5% for binary, 94.95% for three-class and 97.8% for four-class classification [24]. Chauhan et al. used a patch based vision transformer together with a CNN [25], Abdelhaliem et al. used a 3D CNN with adaptive region attention on several MRI modalities [26], and Ravinder et al. used a graph convolutional network [27]. Asiri et al. used adaptive Wiener filtering and independent component analysis for enhancement before an SVM [28]. The price of this progress is complexity: more stages, more parameters, more data and more computing power. Vision transformers, for example, usually need a large and varied dataset to learn well [29].

D. Segmentation

For finding the tumor region, encoder-decoder networks of the U-Net family are the standard [30]. Yan et al. added residual blocks and squeeze-and-excitation blocks to U-Net for multimodal segmentation and improved the Dice score on the BraTS data [31], and Hussain et al. compared U-Net, attention U-Net and residual-attention U-Net for segmentation and classification [32]. Alyami et al. used a deep CNN with the Salp Swarm algorithm for localisation and classification [33]. Such networks need pixel-level masks drawn by experts, which are costly and slow to get. Cheng et al. showed that using the tumor region itself, by augmentation and partition of the region, improves the classification [34], which supports the idea of giving the extracted tumor region to the classifier.

E. Fuzzy methods

Fuzzy logic has been used for two jobs in this area: improving the image and splitting it into regions. Chaki et al. used fuzzy based brightness and contrast improvement [23], and Sowjanya et al. described a fuzzy thresholding technique for dividing a picture into several regions [12]. The fuzzy c-means clustering of Bezdek [11] is the base of many such methods. Its strong point is that it needs no labelled masks and it works with a few lines of code; its weak point is that the result depends on the initial centres, on the number of clusters and on the image quality, which is why we add a brain mask and region rules in our pipeline.

F. Explainable AI in medical imaging

Explainability is now seen as a basic need for medical AI [9], [10]. Class activation based methods are popular for CNNs because they need no change in the network. Grad-CAM [13] uses the gradients of the class score and Grad-CAM++ [35] improves the localisation when several objects of the same class are present. In brain MRI, however, a heatmap can also light up on bright non-tumor structures such as the skull, which makes the explanation less useful. Restricting the heatmap to a segmented region is one simple way to avoid this, and it is what we do in this paper.

G. Review papers

Several surveys exist. Neamah et al. reviewed deep learning models for brain tumor classification and detection [36], Solanki et al. looked at about 100 articles on intelligence techniques [37], and Sumeyye et al. compared 81 studies on classification and grading [38]. The common message is that studies use different datasets, preprocessing and metrics, so the numbers cannot be compared directly, that many methods need high computing power, and that few of them give detection, localisation and explanation together.

H. Summary and research gap

Table IV summarises some of the works above. Looking at them together, the gaps are clear. Most strong classifiers are heavy and give only a label. Segmentation networks need masks. Fuzzy methods are used for enhancement or for segmentation but are rarely linked with a classifier and an explanation. Our work does not try to beat the heavy models in accuracy. The aim is a light pipeline which does detection, localisation, size estimation and explanation together, and to report honestly how well it works.

I. Positioning of this work

Compared with the works above, this paper takes a different position. It does not use a large pre-trained backbone, it does not need expert masks, and it does not claim the best accuracy. It uses an unsupervised fuzzy step to find the tumor, a small network trained from scratch to classify, and a gradient based map to explain, and it reports the weak points openly. We think that this kind of baseline is useful, because it shows what a light and explainable pipeline can do before the extra tricks are added, and it gives a clear point from which later improvements, such as transfer learning or using the tumor region as input, can be measured.

J. Lessons taken from the literature

Reading the works together gives some lessons which shaped this paper. First, enhancement of the image before the network is common and helpful, so we kept a simple pre-processing chain. Second, the tumor region itself carries useful information for classification [34], so the segmentation output is not only a picture but also a possible input to the classifier in the next version. Third, transfer learning is the main reason for the high accuracy of the best papers, so we expect it to close a large part of the gap in our results. Fourth, many papers give accuracy only, and it is hard to see their weak classes, so we report the full confusion matrix. Fifth, explanations are rare, and when they exist they are shown for the whole image; we limit them to the tumor region. These points are not new results of ours; they are the reasons for the choices made in the next section.

TABLE IV. SUMMARY OF SOME RELATED WORKS

Work	Method	Main limitation
Noreen et al.	Inception-V3 + DenseNet201 features	Two networks, no localisation
Majib et al.	VGG16 based CNN	Many parameters, detection only

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

Work	Method	Main limitation
Shah et al.	Fine-tuned EfficientNet	Tumor / non-tumor only
Rizwan et al.	Gaussian CNN	Needs large data, no tumor boundary
Younis et al.	Fine-tuned ResNet50	Deep, no segmentation
Yan et al.	SEResU-Net	Needs masks, no classification
Chaki et al.	Fuzzy enhancement + Inception-ResNet + RL	Many stages, high cost
Asiri et al.	Wiener + ICA + SVM	Hand made pipeline, limited features
Preetha et al.	3B Net + EfficientNetB2	No tumor region shown
Chauhan et al.	Patch ViT + CNN	Data and compute hungry
Abdelhaliem et al.	3D CNN + region attention	Needs several modalities
This work	FCM + small CNN + Grad-CAM	Lower accuracy so far

IV. PROPOSED METHODOLOGY

A. System overview

Fig. 4 shows the steps of the system. The MRI images are read from the dataset and pre-processed. Fuzzy thresholding then extracts the tumor region and the tumor size is calculated. The images are split into training, validation and testing sets and given to the CNN, which predicts one of four classes: glioma, meningioma, pituitary tumor and no tumor. If the result is not good enough the hyperparameters are changed and the network is trained again, otherwise the trained model is saved. At the end Grad-CAM makes the heatmap and the final decision is shown to the doctor with the visual explanation. The segmentation branch and the classification branch work on the same pre-processed image; the branches meet again only at the explanation stage, where the tumor mask is used to clean the Grad-CAM map.

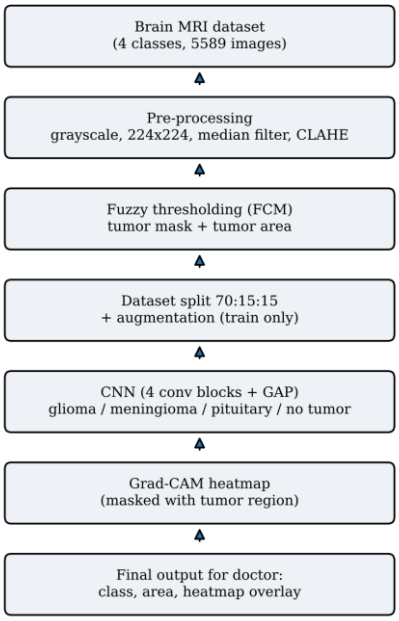


Fig. 4. Block diagram of the proposed system.

B. Dataset

The brain MRI images are loaded into an image datastore from a folder which has one sub-folder for every class, and the label of an image is taken from its folder name. There are four classes, glioma, meningioma, notumor and pituitary. The dataset has 5589 images: 1400 each for glioma, meningioma and notumor, and 1389 for pituitary, so the classes are nearly balanced and no class weighting was needed. The images are of different planes, contrast and head positions (Fig. 5), which makes the task harder, because the same tumor type looks quite different in an axial and in a sagittal slice. For every class the images are divided randomly into training, validation and testing sets in the ratio 70:15:15 (Table V). The training set is used to learn the weights, the validation set to watch the training, and the testing set is kept aside for the unseen images.

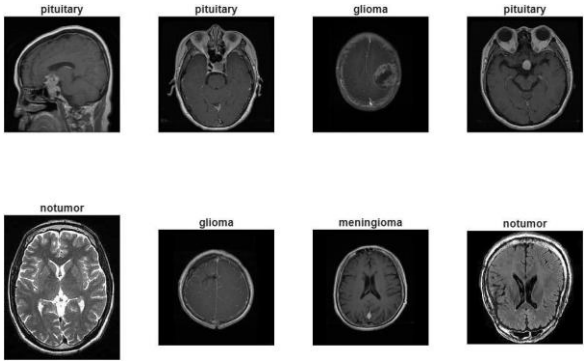


Fig. 5. Sample MRI images from the dataset.

TABLE V. DATASET SPLIT

Subset	Ratio	Images per class	Total images
Training	70%	980	3920
Validation	15%	210	840
Testing	15%	210	840

C. MRI image pre-processing

MRI images differ in size and contrast and they have noise, so every image goes through the following steps when it is read. (i) If it has three channels it is converted to grayscale, because the tumor information is in the intensity and the colour channels of the stored files carry nothing extra. (ii) It is resized to 224 × 224, the input size of the CNN. (iii) A 3 × 3 median filter replaces each pixel by the median of its neighbours; the median filter removes spotty noise and, unlike an averaging filter, it keeps the edges sharp, which matters for the tumor border [39]. (iv) CLAHE improves the local contrast [40]. Ordinary histogram equalisation can over-amplify the noise in flat regions, and CLAHE avoids this by clipping the histogram in small tiles [41]. (v) The gray image is copied into three channels to make a 224 × 224 × 3 image. The input layer of the CNN then scales the pixel values to the range 0 to 1. Pre-processing is done when the image is read, so the stored dataset is never changed.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

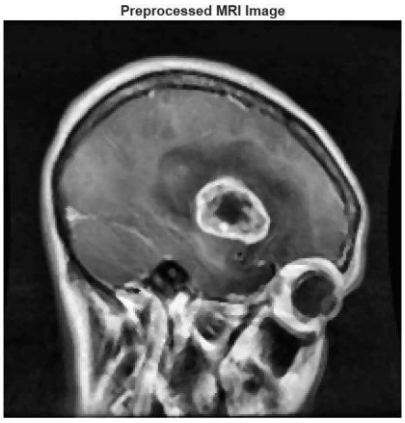


Fig. 6. A glioma MRI image after pre-processing.

D. Fuzzy thresholding and tumor extraction

In normal thresholding each pixel goes fully to one region. In MRI the border between tumor and normal tissue is blurred, so many pixels belong a little to both. In fuzzy thresholding every pixel gets a membership value between 0 and 1 for each region. We use fuzzy c-means (FCM) clustering [11] for this. The pixel intensities are grouped into c = 3 clusters, which roughly correspond to the dark background and fluid, the normal tissue, and the bright structures including the tumor. If x_i is the intensity of pixel i, N the number of pixels, v_j the centre of cluster j, u_{ij} the membership of pixel i in cluster j and m the fuzziness exponent, FCM minimises

$$J = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^{m} \|x_i - v_j\|^2 \qquad (1)$$

The memberships and the centres are updated one after the other with

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{2/(m-1)}} \qquad (2)$$

$$v_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \qquad (3)$$

until the change is very small. We used m = 2, at most 100 iterations and a stopping tolerance of 10^{-5} (Table VI). With m = 2 the membership falls smoothly from the centre of a cluster to its border, and a larger m would make the clusters too fuzzy. The cluster with the highest centre value is taken as the tumor cluster, because the tumor looks brighter than the normal tissue in the images we used. This rule is simple but it is also a weak point, since a dark tumor would be missed (see Section VI).

TABLE VI. SETTINGS OF THE SEGMENTATION STAGE

Setting	Value
Median filter window	3 × 3
FCM clusters (c)	3
Fuzziness exponent (m)	2
Maximum iterations	100
Stopping tolerance	10^{-5}
Tumor membership threshold	0.42
Assumed pixel size	0.26 mm × 0.26 mm

The membership map alone is not enough, because the skull and some other bright structures also have high membership in the brightest cluster. So a mask generation step is added (Algorithm 1 and Fig. 7). The brain is first separated from the dark background with Otsu's threshold [42]. The holes of the brain mask are filled, only the largest connected region is kept, and the mask is eroded so that the skull edge is not included. The membership values of the tumor cluster are then thresholded at 0.42 and only the pixels inside the brain are kept. Very small objects are removed, and the region which lies in the central part of the brain and has a suitable size is chosen as the main tumor region. A morphological opening cuts the thin tails attached to it and the largest remaining part is the final tumor mask. The tumor region image keeps the pixels of the pre-processed image inside this mask and sets all other pixels to zero.

Algorithm 1. Fuzzy tumor extraction

1. Input: pre-processed gray image I (224 × 224).
2. Run FCM on the intensities of I with c = 3, m = 2, 100 iterations, tol = 10^{-5}.
3. Select the cluster with the largest centre as the tumor cluster T.
4. Brain mask B = Otsu(I); fill holes; keep largest region; erode.
5. M = (membership of T > 0.42) AND B.
6. Remove small objects from M.
7. Keep the connected region of M that is central and of valid size.
8. Apply morphological opening; keep the largest component.
9. Output: tumor mask M, tumor region image I·M, area A.

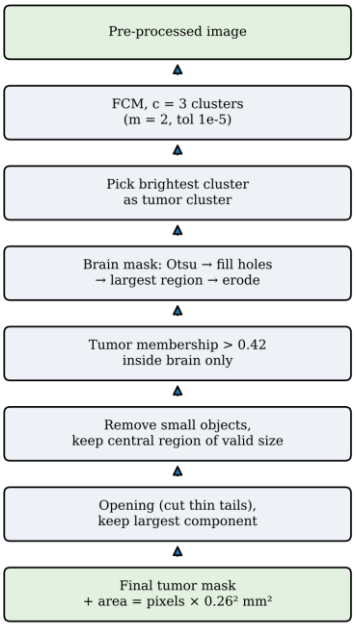


Fig. 7. Steps of the tumor mask generation.

The tumor size is calculated from the number of pixels N_t in the final mask, where each pixel is assumed to cover 0.26 mm × 0.26 mm:

$$A_{tumor} = N_t \times A_{pixel} \qquad (4)$$

For the sample glioma image the mask has 928 pixels, so the estimated area is 928 × 0.0676 = 62.7 mm². The pixel size is an assumption; real scanners store the true pixel spacing in the image header, and using it would make the value clinically meaningful.

E. CNN classifier

The classifier is a CNN designed for this work. It has four convolutional blocks, a global average pooling layer and one

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

fully connected layer (Table VII and Fig. 8). The first three blocks have a convolution, batch normalization [43], ReLU and 2 × 2 max pooling. The fourth block has no max pooling because global average pooling follows it. The number of filters doubles in every block (32, 64, 128, 256) while the feature maps become smaller (224, 112, 56, 28), so the network moves from edges and textures in the early layers to shapes and larger patterns in the last layer. Global average pooling replaces each feature map with its mean [44]; this removes the big fully connected layers which are used in older networks like VGG [45], so the number of parameters is smaller and over-fitting is less likely. The four output scores z are changed into class probabilities by Softmax,

$$P_k = \frac{e^{z_k}}{\sum_{j=1}^{4} e^{z_j}} \qquad (5)$$

and the network is trained by minimising the cross-entropy loss.

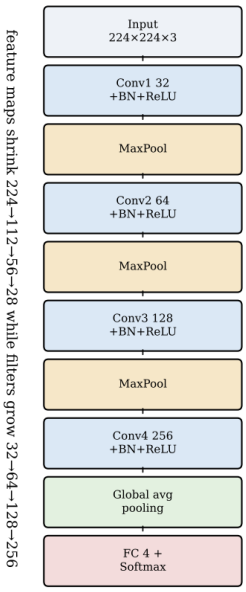


Fig. 8. Layer layout of the proposed CNN.

TABLE VII. ARCHITECTURE OF THE PROPOSED CNN

No.	Layer	Configuration	Output
1	Image input	224×224×3, scaled 0–1	224×224×3
2	Conv1+BN+ReLU	32 filters, 3×3, same	224×224×32
3	Max pool 1	2×2, stride 2	112×112×32
4	Conv2+BN+ReLU	64 filters, 3×3, same	112×112×64
5	Max pool 2	2×2, stride 2	56×56×64
6	Conv3+BN+ReLU	128 filters, 3×3, same	56×56×128
7	Max pool 3	2×2, stride 2	28×28×128
8	Conv4+BN+ReLU	256 filters, 3×3, same	28×28×256
9	Global avg. pooling	mean of each map	1×1×256
10	Fully connected	4 neurons	1×1×4
11	Softmax + class.	class probabilities	4 classes

From the layer sizes we can count the learnable parameters (Table VIII). A convolution with k × k filters, c_in input channels and c_out filters has k·k·c_in·c_out weights plus c_out biases, and a batch normalization layer has two learnable values per channel. In total the network has 390,404 learnable parameters, and about 76% of them are in the last convolutional layer (Fig. 9). This is very small when compared with VGG16 or ResNet50, which have tens of millions, and it means that training is cheap and the model file is small (about 1.5 MB if stored as 32-bit numbers). The price is that the network has less capacity to separate classes which look similar.

TABLE VIII. LEARNABLE PARAMETERS OF THE CNN

Layer	Weights + biases	Batch norm
Conv1 (3→32)	896	64
Conv2 (32→64)	18,496	128
Conv3 (64→128)	73,856	256
Conv4 (128→256)	295,168	512
Fully connected (256→4)	1,028	–
Total	390,404	

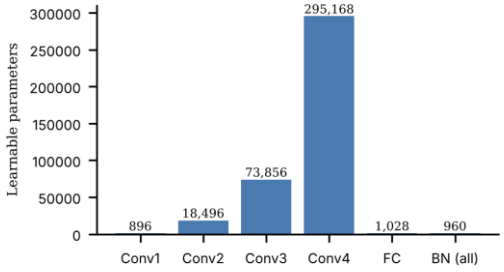


Fig. 9. Learnable parameters per layer.

F. Data augmentation and training

To increase the variety of the training data, the training images are randomly rotated between −10° and +10°, shifted by up to 10 pixels horizontally and vertically, and reflected horizontally. These changes are small, so the image still looks like a real MRI slice. Augmentation is used only for the training images; the validation and testing images are not changed. The network is trained with the Adam optimizer [46] and the options in Table IX. The validation data is used to watch the accuracy and the loss while training, and the trained network is saved in a file so that it can be used directly for the evaluation and for Grad-CAM.

TABLE IX. TRAINING OPTIONS OF THE CNN

Parameter	Value
Optimizer	Adam
Initial learning rate	0.0001
Maximum epochs	3
Mini-batch size	64
Shuffling	Every epoch
Validation frequency	Every 20 iterations

Algorithm 2. Training and explanation

1. Read images, give class labels from folder names.
2. Split every class 70:15:15 into train, validation and test sets.
3. Pre-process each image (gray, 224×224, median, CLAHE, 3 channels).
4. Augment the training images (rotation, shift, flip).
5. Train the CNN with Adam (lr 1e-4, batch 64, 3 epochs).
6. If the result is not sufficient, change hyperparameters and retrain; else save the model.
7. For a new image: predict class c and confidence.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

8. Compute Grad-CAM for c on the last convolutional layer.
9. Multiply the map with the fuzzy tumor mask and overlay it.
10. Show class, area and heatmap to the doctor.

G. Explanation with Grad-CAM

A deep network works like a black box and a doctor cannot rely on a prediction without a reason. Grad-CAM [13] is used to show which part of the image influenced the CNN. The trained CNN first predicts the class c and the confidence (the highest Softmax probability). Then the gradient of the class score y^c with respect to the feature maps A^k of the last convolutional layer is averaged to get a weight for every map,

$$\alpha_k^c = \frac{1}{Z}\sum_{i,j}\frac{\partial y^c}{\partial A_{ij}^k} \qquad (6)$$

where Z is the number of pixels in a feature map. The map is the ReLU of the weighted sum,

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \qquad (7)$$

The ReLU keeps only the features which push the score of the class up. The map is normalised to 0–1, resized to 224 × 224 and multiplied with the fuzzy tumor mask so that the heatmap stays on the tumor. It is shown with the jet colour map on top of the MRI image. Without the mask, a heatmap from the last convolutional layer is coarse (the layer has only 28 × 28 positions) and can spread on the skull; the mask fixes both problems, but it also means that the heatmap cannot show the case when the network looked outside the tumor, so the mask must be seen as a presentation aid and not as a proof of what the network used.

H. Final decision

In the last stage four results are displayed together for the doctor: the input MRI image, the fuzzy tumor mask, the Grad-CAM heatmap and the overlay of the heatmap on the MRI image. The title of the overlay shows the predicted tumor class and the estimated tumor area in mm². In this way the final decision is presented with a visual explanation, and the doctor can compare it with his own reading.

I. Performance measures

The system is evaluated with the confusion matrix, where the actual classes and the predicted classes are compared. For one class, TP are the images of that class classified correctly, FP are images of other classes wrongly classified as this class, and FN are images of this class classified as another class. We calculate

$$Accuracy = \frac{Correctly\ classified\ images}{Total\ images} \qquad (8)$$

$$Precision = \frac{TP}{TP+FP},\quad Recall = \frac{TP}{TP+FN} \qquad (9)$$

$$F1 = \frac{2\times Precision\times Recall}{Precision+Recall} \qquad (10)$$

Recall is also called sensitivity. The class-wise values are averaged over the four classes to get the overall precision, recall and F1-score (macro average). Because the classes have almost the same size, the macro and the size-weighted averages are practically equal here. We also report the specificity of each class, TN / (TN + FP), and Cohen's kappa [47], which tells how much better the classifier is than a random guess that keeps the same class frequencies (0 means chance level and 1 means perfect agreement).

J. Implementation

The complete system is implemented in MATLAB. The pre-processing and the segmentation steps use the Image Processing Toolbox, and the CNN is built, trained and used for Grad-CAM with the Deep Learning Toolbox. The image datastore makes it easy to read the images from the class folders and to apply the pre-processing and the augmentation on the fly.

K. Design choices and alternatives

Some design choices were made on purpose, and the reasons are given here so that the reader can judge them. For the clustering we used FCM and not plain k-means, because k-means gives every pixel to one cluster only and would throw away the partial membership of the border pixels, which is the information that we want. Otsu's global threshold could also split an image into two classes, but it cannot give a membership value, and a single threshold is easily disturbed by the bright skull. We chose c = 3 because the images have three broad intensity groups; with only two clusters the brain tissue and the tumor would often fall in the same group, and with more clusters the tumor ring would be cut into pieces. For the filter we chose the median and not the Gaussian filter because the Gaussian blurs the edges together with the noise. For the network we chose global average pooling in place of flattening the last feature map (28 × 28 × 256 = 200,704 values) into a dense layer, because a dense layer of that size would alone have more weights than the whole present network by a large factor, and our dataset has only about four thousand training images. The learning rate of 0.0001 with Adam is a safe value for a network trained from scratch with batch normalization; it is slow, which is one more reason why three epochs are not enough. We did not test the alternatives against each other in this work, so these are design reasons and not experimental findings.

L. Layer operations in detail

For completeness we write down what the layers of the CNN do. A convolutional layer slides a small filter over the input and, at each position, takes the weighted sum of the covered values and adds a bias:

$$y_{i,j} = \sum_{a,b} w_{a,b}x_{i+a,j+b} + b \qquad (11)$$

Batch normalization then normalises the outputs of the layer with the mean μ and the variance σ² of the mini-batch and rescales them with two learnable values γ and β [43]:

$$\hat{y} = \gamma\frac{y-\mu}{\sqrt{(\sigma^2+\varepsilon)}} + \beta \qquad (12)$$

The ReLU activation keeps positive values and sets negative values to zero, f(x) = max(0, x), which adds the non-linearity and is cheap to compute. The 2 × 2 max pooling keeps the largest value in every 2 × 2 window, so the map becomes half as wide and half as high, and small shifts of the tumor change the output less. Global average pooling takes the mean of each of the 256 maps and gives a vector of 256 values, and the fully connected layer maps this vector to the four class scores. The training minimises the cross-entropy loss over the training images,

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

$$L = -\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{4} t_{ik} log P_{ik} \qquad (13)$$

where M is the number of images in the mini-batch and t_{ik} is 1 if image i belongs to class k and 0 otherwise. Adam updates every weight with its own step size, which is computed from running averages of the gradient and of its square [46], so the learning is less sensitive to the scale of the gradients than plain gradient descent.

M. Walk-through for one image

To make the pipeline concrete, we follow the sample glioma image of Figs. 6, 10 and 15 from the input to the output. The image is read from the glioma folder and its label is taken from the folder name. It is converted to grayscale, resized to 224 × 224, cleaned with the 3 × 3 median filter, enhanced with CLAHE and copied to three channels. In the segmentation branch, FCM groups the intensities into three clusters and the brightest one is taken as the tumor cluster. The brain mask removes the dark background and the skull edge, the tumor memberships above 0.42 inside the brain are kept, and the region selection and the opening leave one compact ring. This mask has 928 pixels, which is 62.7 mm² with the assumed pixel size. In the classification branch, the same image goes through the four convolutional blocks, global average pooling and the Softmax layer, and the class with the highest probability, glioma, is selected. Grad-CAM then takes the gradient of the glioma score with respect to the last convolutional maps, makes the 28 × 28 map, scales it to 224 × 224 and multiplies it with the mask, and the overlay is shown with the title "glioma, area 62.7 mm²". The two branches meet only in the last step.

V. RESULTS AND DISCUSSION

This section presents the results of the proposed method and discusses them. The system was simulated in MATLAB. The results follow the same order as the method: pre-processing, fuzzy segmentation, CNN training, classification performance, error analysis, explanation, and a comparison with the literature.

A. Pre-processing and fuzzy segmentation

After pre-processing, the noise in the glioma image is reduced and the tumor ring and the boundaries of the brain structures are clearly visible (Fig. 6). Fig. 10 shows the result of the fuzzy segmentation for this image. The mask marks the bright ring shaped tumor, and the skull and other bright parts outside the tumor are removed by the brain mask, the region selection and the morphological steps. The mask has 928 pixels, which gives an estimated tumor area of 62.7 mm². The segmentation was shown on this one image, so we cannot yet say how stable the mask is for other images, other planes and other tumor types; this is discussed in Section VI.

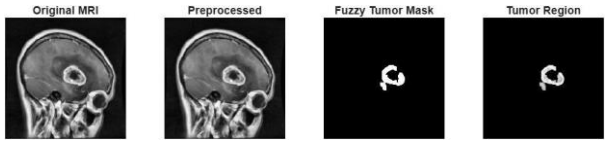


Fig. 10. Fuzzy segmentation result (original, pre-processed, mask, tumor region).

B. CNN training

Fig. 11 gives the training accuracy and loss. The accuracy rises fast in the first epoch and then slowly to about 70–75%, and the loss goes down from about 1.6 to about 0.7. Both curves are still improving a little at the end, so the network has not converged yet. This is expected, because only 3 epochs were used. The curves are also noisy from one iteration to the next, which is normal for small batches. Since the curves are still moving, the final numbers should be seen as a first result and not as the best this network can do.

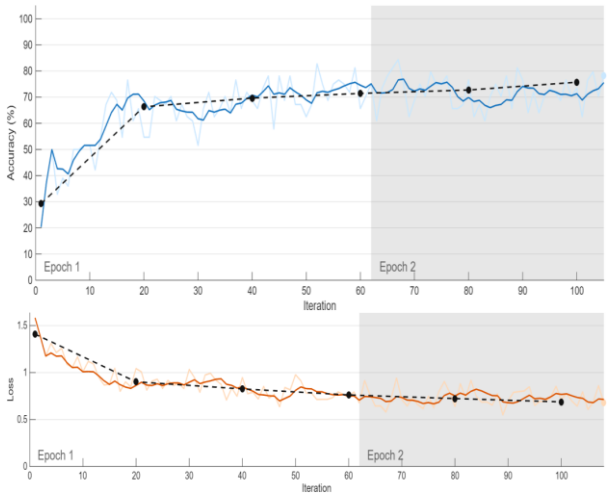


Fig. 11. Training accuracy (top) and training loss (bottom) of the CNN.

C. Classification performance

The confusion matrix is in Fig. 12 (rows are true classes, columns are predicted classes) and the class-wise measures are in Table X and Fig. 13. The matrix contains all 5589 images. The overall accuracy is (1020 + 452 + 1208 + 1340) / 5589 = 4020 / 5589 = 71.93%, and the average precision, recall and F1-score are 75.99%, 71.98% and 70.26%. Cohen's kappa calculated from the same matrix is 0.626, which is a fair agreement with the true labels and clearly above chance, but far from perfect.

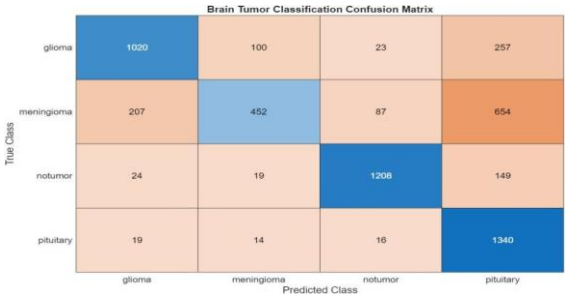


Fig. 12. Confusion matrix of the proposed system.

TABLE X. CLASS-WISE PERFORMANCE (%)

Class	Precision	Recall	F1-score	Sensitivity
Glioma	80.31	72.86	76.40	72.86
Meningioma	77.26	32.29	45.54	32.29
Notumor	90.55	86.29	88.37	86.29
Pituitary	55.83	96.47	70.73	96.47
Average	75.99	71.98	70.26	71.98

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

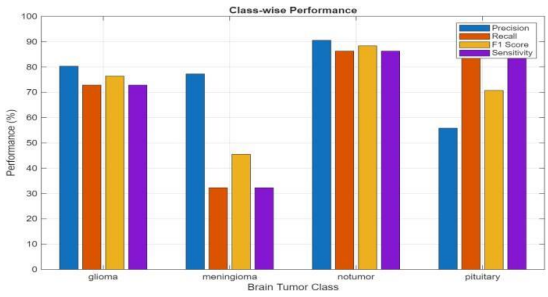


Fig. 13. Class-wise precision, recall, F1-score and sensitivity.

D. Error analysis

The confusion matrix shows more than the accuracy. In total 1569 of the 5589 images are classified wrongly. Of these, 1060 (67.6%) are wrongly sent to the pituitary class, and the single biggest error is meningioma predicted as pituitary, which has 654 images and makes 41.7% of all errors. So the classifier does not fail in an even way; it has one strong bias, towards pituitary. Table XI gives the row-normalised matrix, in which each row is the true class and sums to 100%, and Fig. 14 shows the same as a heatmap. Table XII gives the specificity and the counts for each class, and Figs. 16 and 17 show where the images of each class go and how many false positives and false negatives each class has.

TABLE XI. ROW-NORMALISED CONFUSION MATRIX (%)

True \ Pred.	Glioma	Meningioma	Notumor	Pituitary
Glioma	72.9	7.1	1.6	18.4
Meningioma	14.8	32.3	6.2	46.7
Notumor	1.7	1.4	86.3	10.6
Pituitary	1.4	1.0	1.2	96.5

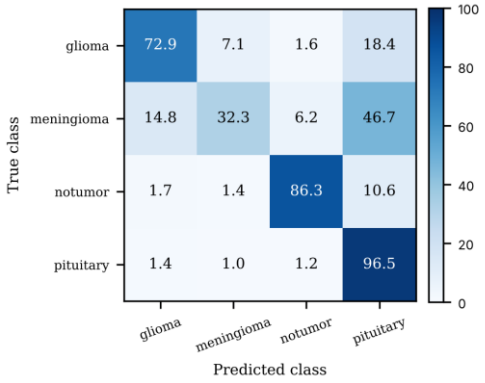


Fig. 14. Row-normalised confusion matrix (percent of each true class).

TABLE XII. PER-CLASS COUNTS AND SPECIFICITY

Class	TP	FP	FN	TN	Specificity (%)
Glioma	1020	250	380	3939	94.03
Meningioma	452	133	948	4056	96.83
Notumor	1208	126	192	4063	96.99
Pituitary	1340	1060	49	3140	74.76

Two classes show the problem clearly. Pituitary has only 49 false negatives but 1060 false positives, so its recall is very high and its precision low, and its specificity (74.76%) is the lowest of the four. Meningioma is the opposite: only 133 false positives but 948 false negatives, so it is almost never predicted wrongly, but it is also missed most of the time. The two classes take each other's place. Meningioma and pituitary tumors both lie near the skull base and look similar in many slices, so they are the most confused pair, and this pulls down their F1-scores (45.54% and 70.73%). Glioma is sent to pituitary in 18.4% of the cases and to meningioma in 7.1%.

E. Tumor versus no tumor

In practice, the first question for a doctor is often whether there is a tumor at all. We can answer this from the same confusion matrix by merging the three tumor classes. There are 4189 tumor images and 1400 no-tumor images. Of the tumor images, only 126 (3.0%) are predicted as notumor, so the tumor detection rate is 97.0%. Of the no-tumor images, 192 (13.7%) are predicted as some tumor class, so the specificity is 86.3%. The merged accuracy is (4063 + 1208) / 5589 = 94.31%, and the precision of a tumor alarm is 95.5%. This view must be read with care, because the network was not trained for this two-class task, and it comes from the same all-image matrix; but it shows that the network has learned to separate normal from abnormal images quite well, and that most of the weakness lies in telling the tumor types apart. Among the 4063 tumor images that are detected as tumor, the three-way type accuracy is only 69.2%, which agrees with the confusion described above.

F. Grad-CAM and final decision

Fig. 15 shows the final output for the glioma test image. The first image is the input MRI, the second is the fuzzy tumor mask, the third is the Grad-CAM heatmap and the last is the overlay of the heatmap on the MRI. The heatmap sits on the ring shaped tumor, and the title of the last image gives the class as glioma with an area of 62.7 mm². With this output the doctor can check by eye whether the region used by the system for its decision is the tumor region. If the class were wrong but the heatmap was still on the tumor, the doctor would know that the problem is in the classification and not in the localisation, and if the heatmap was on a wrong place, the doctor would know not to trust the result.

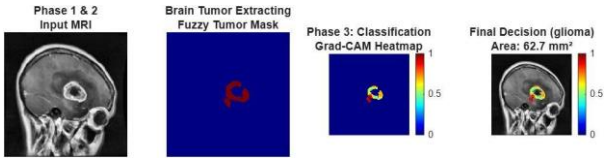


Fig. 15. Final decision with visual explanation (input, mask, Grad-CAM, overlay).

G. Comparison with related work

The reviewed papers report higher accuracy than ours: for example about 97–100% per class for the fuzzy plus Inception-ResNet reinforcement learning model [23] and 94.95–99.5% for the hybrid 3B Net and EfficientNetB2 model [24]. A direct comparison is not fair, because the datasets, splits, pre-processing and the number of training epochs are different, and some of those papers use pre-trained networks with tens of millions of parameters. Table XIII therefore compares the properties of the methods and not only the accuracy. Our network is much smaller, it gives the tumor region and size, and it gives an explanation, but its accuracy is lower, and we do not claim that it is better than the heavy models. The result

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

should be seen as the baseline of a light and explainable pipeline.

TABLE XIII. PROPERTIES OF SOME METHODS

Method	Model size	Tumor region	Explanation	Reported accuracy
Fuzzy + Inc-ResNet + RL [Chaki]	Large	No	No	94–100% per class
3B Net + EffNetB2 [Preetha]	Large	No	No	94.95–99.5%
Fine-tuned EffNet [Shah]	Medium	No	No	Not compared
SEResU-Net [Yan]	Large	Yes	No	Dice only
This work	0.39 M param.	Yes (mask, area)	Yes (Grad-CAM)	71.93%

H. Why the accuracy is not higher

We think that several things together keep the accuracy down. First, the training was short: only 3 epochs, and the curves in Fig. 11 are still going up. Second, the network is small and trained from random weights, without any pre-trained knowledge, while the best papers use pre-trained backbones. Third, the dataset has different planes (axial, sagittal, coronal), different contrast and different head positions mixed together, so the same tumor type has many looks, and a shallow network with 28 × 28 features at the end may not be able to cover all of them. Fourth, small tumors near the skull base (meningioma and pituitary) occupy a few pixels of a 224 × 224 image after resizing, so their features are weak. Fifth, the segmentation output is not used by the classifier at all, so the network cannot profit from the tumor region found by the fuzzy stage. These are guesses supported by the curves and by the error pattern, and we did not run ablation experiments to prove them; the future work below is designed to test them one by one.

I. Computational considerations

Because the CNN has only 0.39 million parameters and uses 3 × 3 convolutions with global average pooling, both training and use are cheap. From the layer sizes the cost can be counted exactly (Table XIV): the three last convolutional layers cost the same, about 231 million multiply-accumulate operations each, because every time the feature map gets four times smaller the number of filters doubles and the number of input channels doubles too, while the first layer is the cheapest. Most of the parameters, however, are in the last layer. The fuzzy c-means step runs on the intensities of one gray image and is also light, since it clusters scalar values. The Grad-CAM step needs only one forward and one backward pass through the network for each image. This makes the whole pipeline suitable for a normal computer without a special GPU, and it is the reason why the FPGA idea in Section VIII is realistic. We did not measure the run time on a specific machine in this work, so we give no timing numbers.

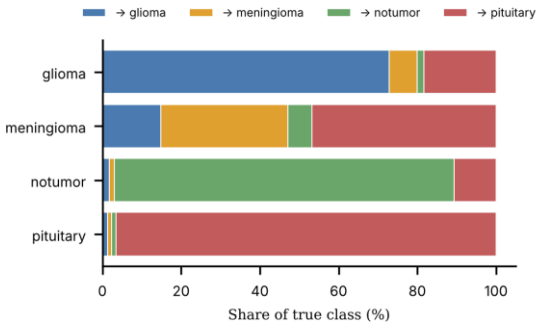


Fig. 16. Where the images of each true class go (percent of the class).

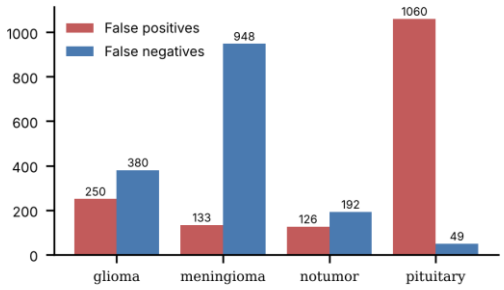


Fig. 17. False positives and false negatives of each class.

TABLE XIV. COMPUTATION AND MEMORY OF THE CONVOLUTIONAL LAYERS

Layer	Map size	Multiply-adds (M)	Output values
Conv1	224×224	43.4	1,605,632
Conv2	112×112	231.2	802,816
Conv3	56×56	231.2	401,408
Conv4	28×28	231.2	200,704
Total (with FC)	–	≈ 737	–

In total one image needs about 737 million multiply-accumulate operations, which is about 1.5 GFLOPs, for the forward pass. For comparison, VGG16 needs roughly 15 billion multiply-accumulate operations and has about 138 million parameters, and ResNet50 needs roughly 4 billion with about 25 million parameters [45], [48] (approximate values from the usual 224 × 224 setting). The proposed CNN is therefore about 5 to 20 times cheaper in computation and about 65 to 350 times smaller in parameters than these networks, which fits the aim of a light pipeline. The memory for the activations is dominated by the first layer (1.6 million values for one image), and the last layer has the smallest map, which is the reason why Grad-CAM from this layer is coarse. Note that these are counts from the architecture and not measured run times.

J. Worked example of the measures

To show how the numbers in Table X come from the matrix, we calculate two classes by hand. For meningioma, the diagonal element is TP = 452. The column of meningioma holds 100 + 452 + 19 + 14 = 585 images, so FP = 585 − 452 = 133 and the precision is 452 / 585 = 77.26%. The row of meningioma holds 207 + 452 + 87 + 654 = 1400 images, so FN = 1400 − 452 = 948 and the recall is 452 / 1400 = 32.29%. The

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

F1-score is 2 × 0.7726 × 0.3229 / (0.7726 + 0.3229) = 45.54%. For pituitary, TP = 1340, the column holds 257 + 654 + 149 + 1340 = 2400 images so the precision is 1340 / 2400 = 55.83%, and the row holds 19 + 14 + 16 + 1340 = 1389 images so the recall is 1340 / 1389 = 96.47% and the F1-score is 70.73%. The two examples show how a class can have a high precision with a low recall (meningioma) or the reverse (pituitary), and why the F1-score is a better single number than accuracy for judging them.

K. Clinical reading of the class-wise results

The different errors do not have the same cost in practice. A tumor which is classified as notumor (126 of 4189 tumor images in our matrix) is the most dangerous case, because the patient could be sent home. This happens rarely here, which is a good sign, although it is not zero. A tumor of one type classified as another type is less dangerous at the screening stage, since the patient still gets further examination, but it can still lead to a wrong plan, because meningioma and glioma are treated in different ways. A notumor image classified as a tumor (192 of 1400) causes a false alarm and extra tests, which is a cost, but a smaller one. With these costs in mind, the high recall of pituitary (96.47%) is useful only if its low precision (55.83%) is accepted, and the low recall of meningioma (32.29%) is the main point which has to be improved before the system could assist in real work. We suggest that future versions report the sensitivity for the tumor versus no-tumor decision separately from the type accuracy, as we did in Section V-E.

L. Expected sensitivity of the pipeline to its settings

Some settings of the pipeline are likely to change the results, and we describe the expected direction here as a guide for later experiments; we did not run these tests. If the membership threshold of 0.42 is raised, the mask would shrink and become cleaner but small tumors could vanish; if it is lowered, the mask would grow and include more of the normal tissue around the tumor, and the area estimate would grow with it. A larger median window would remove more noise but also round off the tumor border. A stronger CLAHE clip limit would increase the local contrast and with it the noise. For the CNN, more epochs and a decaying learning rate are expected to help most, because the training curves are still rising, while a deeper network or pre-trained weights are expected to help with the hard pair of classes. A study which changes one of these settings at a time, and reports the effect with a held-out test set and repeated runs, would be a good next step.

M. Reproducibility

All the settings needed to repeat the work are given in the tables of Section IV: the image size, the filter and the CLAHE step, the FCM settings and the thresholds, the layer sizes of the CNN, the augmentation ranges and the training options. What is not fixed in the description is the random seed used for the 70:15:15 split and for the weight initialisation, so a new run will give somewhat different numbers, mainly because only three epochs are used. For a fair report, the work should be repeated with several seeds and the mean and the spread of the accuracy should be given.

N. Statistical uncertainty

A single accuracy number hides its uncertainty. If the 5589 images are treated as independent samples, the 95% Wilson confidence interval of the overall accuracy is 70.73% to 73.09%, which is about ±1.2 points. The intervals of the class recalls are wider because every class has only about 1400 images (Table XV). The intervals do not overlap for meningioma (29.9–34.8%) and the other classes, so the weakness of meningioma is not a matter of chance. These intervals describe only the sampling noise on this image set. They do not include the effect of the random split and the random initialisation of the network, of the same patient appearing in several images, or of the change to another dataset, so the real uncertainty is larger.

TABLE XV. 95% CONFIDENCE INTERVALS (WILSON, %)

Measure	Value	95% interval	Images
Overall accuracy	71.93	70.73 – 73.09	5589
Glioma recall	72.86	70.47 – 75.12	1400
Meningioma recall	32.29	29.89 – 34.78	1400
Notumor recall	86.29	84.38 – 87.99	1400
Pituitary recall	96.47	95.37 – 97.32	1389

O. How to read the Grad-CAM output

The heatmap uses the jet colour map, in which blue means a value near 0 and red means a value near 1, so the red and yellow parts of the overlay are the places that pushed the score of the predicted class up the most. A doctor can use the overlay in three simple ways. If the red region covers the tumor ring, the network probably used the right evidence. If the red region is small compared with the mask, the network used only a part of the tumor. If the class is doubtful, for example a low confidence value, the doctor should look at the original slices more carefully. Because the map comes from a layer with only 28 × 28 positions and is scaled up to 224 × 224, its borders are smooth and not exact, and it should not be used to measure the tumor; the mask and the area value are for that. The two outputs have different jobs: the mask gives the where and how big, and the heatmap gives the why.

P. Discussion of glioma and no-tumor results

Two classes deserve a short note. The glioma class is classified correctly in 72.9% of the cases (1020 of 1400). Its errors go mainly to pituitary (257 images, 18.4%) and to meningioma (100 images, 7.1%), and only 23 glioma images (1.6%) are called notumor, so glioma is rarely missed completely. This agrees with the fact that gliomas are often large and have a clear ring or irregular shape (like the sample in Fig. 15), which is easier to see than a small tumor at the skull base. The notumor class has 1208 of 1400 correct (86.3%). Its errors are mostly images called pituitary (149, 10.6%), while only 24 are called glioma and 19 meningioma. Looking at the sample images in Fig. 5, some normal slices are taken at the skull base, where the pituitary gland and nearby structures are normally bright, and it is likely that the network learned to connect these bright structures with the pituitary class. We have not checked this explanation image by image, but it fits both the large number of pituitary false positives (1060) and the position of the errors in the matrix.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

VI. LIMITATIONS AND THREATS TO VALIDITY

There are some limits in this work that we want to state clearly.

- **Evaluation set:** the confusion matrix is calculated on all 5589 images and not only on the 840 held-out test images, so it includes images that the network saw during training. The accuracy should therefore not be read as a clean test result, and a true test accuracy could be lower. A separate evaluation on the test split is the first thing to do next.
- **Short training:** only 3 epochs were used, and the curves have not converged, so the numbers are a first result.
- **One dataset:** all images come from one collection, so we do not know how the system behaves on images from another hospital or scanner.
- **Pixel size:** the tumor area uses an assumed pixel size of 0.26 mm × 0.26 mm, so it is only an estimate and not a calibrated clinical value.
- **Segmentation:** the settings (membership threshold 0.42, size limits of the region, the rule that the brightest cluster is the tumor) are fixed numbers chosen for this dataset, and the segmentation was demonstrated on one image only. No overlap score such as Dice could be given because no ground-truth masks were used. Dark or very small tumors may be missed.
- **Explanation:** because the heatmap is multiplied with the tumor mask, it always looks concentrated on the tumor, so it cannot reveal a case where the network used something else.
- **Slices, not patients:** the images are 2-D slices. If slices from one patient are in both the training and the test set, the result can be too optimistic. The dataset description available to us does not say how this was handled.
- **Not a diagnosis:** the system is a support tool and its output is not a medical diagnosis.

VII. ETHICAL AND CLINICAL CONSIDERATIONS

A tool like this touches the health of people, so some points need care. The first one is the danger of wrong answers. In our results meningioma is missed in two cases out of three, and a doctor who trusts the label blindly could make a wrong decision. For this reason the system shows the heatmap and the mask, so that a human can check the result, and the output should always be treated as a second opinion. The second point is the data. Medical images are private, and any real use needs the consent of the patients, removal of identity information and the approval of the hospital. The third point is fairness: if the training images come from one group of patients or one type of scanner, the system may work worse for other groups, and this must be tested before any use. The last point is responsibility: the final decision must stay with the clinician, and the system should be tested and validated in a proper clinical study before it is used in practice [10].

VIII. CONCLUSION AND FUTURE WORK

A. Conclusion

This paper presented a brain tumor detection and classification system which joins pre-processing, fuzzy c-means thresholding, a four-class CNN and Grad-CAM. The fuzzy stage gave a clear tumor mask and size, and Grad-CAM gave a visual explanation of the decision. The CNN, which has about 0.39 million parameters, reached 71.93% accuracy, with an average precision of 75.99%, recall of 71.98% and F1-score of 70.26%. The notumor class was best (F1 88.37%) and meningioma was weakest (recall 32.29%) because many of its images were classified as pituitary. Most of the 1569 errors come from this pair of classes, and when the tumor classes are merged the same predictions give 94.31% accuracy for tumor versus no tumor. So the pipeline already separates normal from abnormal images fairly well, but it cannot yet tell the tumor types apart with enough reliability. The reasons are mainly the short training, the small network and the mixed image planes, and the evaluation also needs to be repeated on held-out images only.

B. Summary of findings

The main findings can be summarised in five points. (1) The fuzzy thresholding pipeline gave a clean tumor mask and an area estimate on the sample image. (2) The small CNN with 0.39 million parameters reached 71.93% four-class accuracy after only three epochs, with a kappa of 0.626. (3) The errors are not spread evenly: 67.6% of them are images wrongly sent to pituitary, and 41.7% are meningioma images predicted as pituitary. (4) The same predictions give 94.31% accuracy and a 97.0% detection rate when the question is only tumor or no tumor. (5) The Grad-CAM heatmap restricted by the mask gives a clear visual explanation, but it is a presentation aid and cannot show when the network looked at a wrong place.

C. Future work

In the future we plan to do the following. (1) Train for more epochs and evaluate only on the held-out test images, with a patient-wise split if the information is available. (2) Use transfer learning with pre-trained networks such as ResNet [48] or EfficientNet [49]. (3) Try class-weighted loss and stronger augmentation to reduce the meningioma–pituitary confusion. (4) Give the extracted tumor region to the CNN as an extra input, because the region alone has been shown to help classification [34]. (5) Measure the segmentation with Dice and IoU on images with expert masks, and read the true pixel size from the image header. (6) Test the system on images from other hospitals, together with clinical centres. (7) Deploy the system as a web application for real-time brain tumor classification.

A hardware direction is also planned. The proposed system is expected to fit on a low-cost FPGA of the Spartan-3 Xilinx family, in about 300 slices, which corresponds to roughly 5,450 equivalent gates. The expected clock is 80 MHz and the expected processing time is 0.339 ms per image, which would allow about 3,000 images per second (Table XVI). These are design estimates and not measured results. Another direction is a system that extracts and shows each tumor type separately inside the MRI image, which would make the output easier to read, and a version that works with several MRI sequences.

TABLE XVI. EXPECTED FPGA IMPLEMENTATION (ESTIMATES)

Item	Expected value
Target device	Spartan-3 (Xilinx), low cost
Area	≈ 300 slices
Equivalent gates	≈ 5,450
Clock frequency	80 MHz

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

Item	Expected value
Time per image	0.339 ms
Throughput	≈ 3,000 images/s

D. Possible deployment workflow

If the system is improved enough, it could be used as a support tool in the way shown in Table XVII. The radiologist loads the slices, the system pre-processes them, shows the mask, the class, the area and the heatmap, and the radiologist accepts, corrects or rejects the result. The corrections can be saved and later used to retrain the network, so the system improves with use, but only after the hospital has approved the process.

TABLE XVII. POSSIBLE WORKFLOW IN A CLINIC

Step	Who / what	Action
1	Radiology staff	Load MRI slices of the patient into the tool
2	System	Pre-process, extract mask, estimate area
3	System	Predict class and confidence, make heatmap
4	Radiologist	Check class, mask and heatmap against own reading
5	Radiologist	Accept, correct or reject; decide next examination
6	Hospital	Store corrected cases for later validation and retraining

APPENDIX A. ABBREVIATIONS

Short	Meaning
AI	Artificial Intelligence
BN	Batch Normalization
CAM	Class Activation Mapping
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
CT	Computed Tomography
DL / ML	Deep Learning / Machine Learning
FCM	Fuzzy C-Means
FN / FP	False Negative / False Positive
Grad-CAM	Gradient-weighted Class Activation Mapping
MRI	Magnetic Resonance Imaging
ReLU	Rectified Linear Unit
TP	True Positive
XAI	Explainable Artificial Intelligence

APPENDIX B. NOTATION

Symbol	Meaning
N	Number of pixels in the image
c	Number of FCM clusters (3)
m	Fuzziness exponent (2)
x_i	Intensity of pixel i
v_j	Centre of cluster j
u_{ij}	Membership of pixel i in cluster j
N_t	Number of pixels in the tumor mask
z_k, P_k	Output score and Softmax probability of class k
y^c	Score of class c before Softmax
A^k	k-th feature map of the last convolutional layer
α_k^c	Grad-CAM weight of feature map k for class c
Z	Number of pixels in a feature map

REFERENCES

[1] G. D. Barkade, P. S. Bhosale, and S. K. Shirsath, “Overview of brain cancer, its symptoms, diagnosis and treatment,” IP Int. J. Comprehensive Adv. Pharmacol., vol. 8, no. 3, pp. 159–164, Nov. 2023.

[2] D. N. Louis et al., “The 2021 WHO classification of tumors of the central nervous system: A summary,” Neuro-Oncology, vol. 23, no. 8, pp. 1231–1251, Aug. 2021.

[3] M. Sasikala and N. Kumaravel, “A wavelet-based optimal texture feature set for classification of brain tumours,” J. Med. Eng. Technol., vol. 32, no. 3, pp. 198–205, Jan. 2008.

[4] E.-S. A. El-Dahshan, T. Hosny, and A.-B. M. Salem, “Hybrid intelligent techniques for MRI brain images classification,” Digit. Signal Process., vol. 20, no. 2, pp. 433–441, Mar. 2010.

[5] K. Kaplan, Y. Kaya, M. Kuncan, and H. M. Ertunç, “Brain tumor classification using modified local binary patterns (LBP) feature extraction methods,” Med. Hypotheses, vol. 139, Art. no. 109696, Jun. 2020.

[6] C. S. Rao and K. Karunakara, “Efficient detection and classification of brain tumor using kernel based SVM for MRI,” Multimedia Tools Appl., vol. 81, no. 5, pp. 7393–7417, Feb. 2022.

[7] M. M. Badža and M. Č. Barjaktarović, “Classification of brain tumors from MRI images using a convolutional neural network,” Appl. Sci., vol. 10, no. 6, p. 1999, Mar. 2020.

[8] M. Hammad, M. ElAffendi, A. A. Ateya, and A. A. A. El-Latif, “Efficient brain tumor detection with lightweight end-to-end deep learning model,” Cancers, vol. 15, no. 10, p. 2837, May 2023.

[9] F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” arXiv:1702.08608, 2017.

[10] E. Tjoa and C. Guan, “A survey on explainable artificial intelligence (XAI): Toward medical XAI,” IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 11, pp. 4793–4813, Nov. 2021.

[11] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms. New York, NY, USA: Plenum, 1981.

[12] K. Sowjanya, M. Ajreen, P. Sidharth, K. Sriharsha, and L. A. Rao, “Fuzzy thresholding technique for multiregion picture division,” Int. Res. J. Adv. Sci. Hub, vol. 4, no. 3, pp. 45–50, Mar. 2022.

[13] R. R. Selvaraju et al., “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 618–626.

[14] A. Pashaei, H. Sajedi, and N. Jazayeri, “Brain tumor classification via convolutional neural network and extreme learning machines,” in Proc. 8th Int. Conf. Comput. Knowl. Eng. (ICCKE), Mashhad, Iran, Oct. 2018, pp. 314–319.

[15] Z. Rasheed et al., “Brain tumor classification from MRI using image enhancement and convolutional neural network techniques,” Brain Sci., vol. 13, no. 9, p. 1320, Sep. 2023.

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

[17] M. F. I. Soumik and M. A. Hossain, “Brain tumor classification with inception network based deep learning model using transfer learning,” in Proc. IEEE Region 10 Symp. (TENSYMP), Jun. 2020, pp. 1018–1021.

[18] N. Noreen et al., “A deep learning model based on concatenation approach for the diagnosis of brain tumor,” 2020.

[19] M. S. Majib et al., “VGG-SCNet: A VGG net-based deep learning framework for brain tumor detection on MRI images,” 2021.

[20] H. A. Shah et al., “A robust approach for brain tumor detection in magnetic resonance images using finetuned EfficientNet,” 2022.

[21] A. Younis et al., “Abnormal brain tumors classification using ResNet50 and its comprehensive evaluation,” 2024.

[22] M. Rizwan et al., “Brain tumor and glioma grade classification using Gaussian convolutional neural network,” 2022.

[23] J. Chaki et al., “Brain tumor categorization and retrieval using deep brain incep res architecture based reinforcement learning network,” 2023.

[24] R. Preetha et al., “Hybrid 3B Net and EfficientNetB2 model for multi-class brain tumor classification,” 2025.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

[25] P. Chauhan et al., “PBViT: A patch-based vision transformer for enhanced brain tumor detection,” 2025.

[26] I. Abdelhaliem et al., “A multimodal adaptive inter-region attention-guided network for brain tumor classification,” 2025.

[27] M. Ravinder et al., “Enhanced brain tumor classification using graph convolutional neural network architecture,” Sci. Rep., vol. 13, no. 1, pp. 1–22, Sep. 2023.

[28] A. A. Asiri et al., “Optimized brain tumor detection: A dual-module approach for MRI image enhancement and tumor classification,” 2024.

[29] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” in Proc. ICLR, 2021.

[30] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. MICCAI, 2015, pp. 234–241.

[31] C. Yan et al., “SEResU-Net for multimodal brain tumor segmentation,” 2022.

[32] S. S. Hussain et al., “Next-generation automation in neuro-oncology: Advanced neural networks for MRI-based brain tumor segmentation and classification,” 2025.

[33] J. Alyami et al., “Tumor localization and classification from MRI of brain using deep convolution neural network and Salp Swarm algorithm,” Cognit. Comput., vol. 16, no. 4, pp. 2036–2046, Jul. 2024.

[34] J. Cheng et al., “Enhanced performance of brain tumor classification via tumor region augmentation and partition,” PLoS ONE, vol. 10, no. 10, Art. no. e0140381, Oct. 2015.

[35] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, “Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks,” in Proc. IEEE WACV, 2018, pp. 839–847.

[36] K. Neamah et al., “Brain tumor classification and detection based DL models: A systematic review,” 2024.

[37] S. Solanki et al., “Brain tumor detection and classification using intelligence techniques: An overview,” 2024.

[38] Sumeyye et al., “A systematic review of machine learning and deep learning techniques for brain tumor classification, and grading,” 2026.

[39] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 4th ed. New York, NY, USA: Pearson, 2018.

[40] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in Graphics Gems IV, P. S. Heckbert, Ed. San Diego, CA, USA: Academic Press, 1994, pp. 474–485.

[41] S. M. Pizer et al., “Adaptive histogram equalization and its variations,” Comput. Vis., Graph., Image Process., vol. 39, no. 3, pp. 355–368, 1987.

[42] N. Otsu, “A threshold selection method from gray-level histograms,” IEEE Trans. Syst., Man, Cybern., vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[43] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in Proc. ICML, 2015.

[44] M. Lin, Q. Chen, and S. Yan, “Network in network,” in Proc. ICLR, 2014.

[45] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. ICLR, 2015.

[46] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. ICLR, 2015.

[47] J. Cohen, “A coefficient of agreement for nominal scales,” Educ. Psychol. Meas., vol. 20, no. 1, pp. 37–46, 1960.

[48] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.

[49] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. ICML, 2019.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)