Journal of Theoretical and Applied Information Technology
15th February 2025. Vol.103. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# DEEP LEARNING APPROCHES FOR TUMOR DETECTION USING MRI DATA

**PURNACHANDRARAO MURALA[1], KUNJAM NAGESWARA RAO[2]**

[1]Research Scholar, Department of CS&SE, Andhra University, Visakhapatnam, India

[2]Professor, Department of CS&SE, Andhra University, Visakhapatnam, India

E-mail: [1]purnachandrarao.m@gmail.com, [2]kunjamnag@gmail.com

**ABSTRACT**

Neurological diseases are relatively severe in the field of health informatics, often associated with life-threatening symptoms and costly treatments. Among these, brain tumors stand out as a well-known concern, showing a noteworthy increase in the affected patient's number over the past decade. MRI imaging is the primary method for tumor detection, and recent advancements in Computer-Aided Diagnosis (CAD) using deep learning have improved diagnostic accuracy. However, existing models have drawbacks, such as inadequate dataset sizes, which hinder early-stage tumor detection, and the limited number of extracted features from input images. To address the challenges, an Enhanced Convolutional Neural Network (ECNN) model has been proposed. The proposed ECNN is trained on the MRI images taken from the BR35H benchmark dataset with extensive data augmentation techniques to improve generalization. The ECNN model achieved a high accuracy of 99.3% in classifying tumor images. Once tumor-positive images were identified, further analysis was performed using a Vision Transformer (ViT) model, trained on a different subset of the BR35H dataset. The ViT model achieved an accuracy of 97% in localizing tumor regions, showing its effectiveness in precise tumor segmentation. This hybrid technique, which addresses important issues in automated brain tumor diagnosis, improves both detection accuracy and interpretability by using CNN-based classification followed by Transformer-based localization.

**Keywords**: *Brain Tumors, MRI Images, Computer-Aided Diagnosis (CAD), Enhanced Convolutional Neural Network (ECNN)s, Vision Transformer (VIT).*

## 1. INTRODUCTION

Brain tumors pose a significant health challenge, impacting lives throughout the world. Though medical research and technology are advanced, the incidence of brain tumors is continuously increasing and affecting people of all age groups. A brain tumor refers to an abnormal mass of cells caused by uncontrolled cell division within the brain. If appropriate measures are not taken at an early stage, these groups of cells harm healthy cells, which reduces the functionality of the brain [1]. Computer-aided diagnosis (CAD) helps neuro-oncologists in different ways to detect brain tumors [3]. CAD includes diagnostic systems that are based on deep learning and machine learning [2]. The increasing use of CAD for tumor detection is largely attributed to advancements in medical imaging technologies like MRI and CT scans, coupled with digital image processing techniques [4].

Magnetic Resonance Imaging (MRI) is the most commonly and widely used option for diagnosing brain tumors. MRI provides high-resolution images of the brain, offering detailed contrast in soft tissues, which facilitates the detection of abnormalities. Traditionally, radiologists manually analyze MRI scans to identify irregularities in the brain. However, with the rapid advancements in medical imaging hardware and procedures, the volume of MRI data has grown exponentially, making manual analysis challenging and time-consuming. This has created a need for automation or semi-automation of the manual procedures, where machine learning plays an important role. Machine learning has become the powerful tool for improving performance in different medical applications in different fields, such as identifying molecular and cellular structures, tissue segmentation, and classification of images [5-7].

Deep learning, which is a part of machine learning, has become an innovative tool in medical imaging that provides unmatched precision and

efficiency in diagnostic tasks. It involves multi-layered neural networks that can learn hierarchical features directly from raw data. Among the several deep learning architectures, CNNs are effective for the image-based applications. CNNs are good at extracting spatial and contextual features from medical images, making them a better option for brain tumor detection. By applying different filters over MRI images, CNNs can detect patterns that are essential for differentiating between tumor and non-tumor regions. For tasks like tumor classification and segmentation, advanced CNN-based models are being widely utilized in CAD systems to improve accuracy over traditional approaches.

In addition to CNNs, specialized architectures such as U-Net and ViTs are also receiving attention in medical imaging for tasks that need accurate localization and segmentation. U-Net, a popular encoder-decoder network, is well known for biomedical image segmentation because of its symmetric structure and skip connections, which ensure detailed spatial information remains accessible throughout the network. Even with small training data, U-Nets can precisely define the borders of tumors in MRI scans. On the other side, ViTs use self-attention mechanisms to identify long-range dependencies within images, making them effective for both segmentation and classification tasks. By processing images as patches and concentrating on inter-patch interactions, ViTs are able to identify small abnormalities in MRI images, where traditional models might miss.

## 2. RELATED WORK

[7] Saeedi, Rezayi, and Keshavarz employed a combination of 2D CNN and autoencoder networks, along with machine learning techniques like MLP and KNN, for detecting brain tumors from MRI images. The 2D CNN achieved a peak accuracy of 96.47% on a dataset of 3,264 MRI images, demonstrating the effectiveness of integrating unsupervised learning with traditional classification methods to enhance feature extraction.

[8] Srikanth B. and Suryanarayana S.V. introduced a method for multi-class brain tumor classification, combining data augmentation with a deep neural network. They used a 16-layer VGG-16 model, leveraging advanced feature extraction methods to achieve an accuracy of 98% on a dataset collected in China from 2010 to 2015. This approach illustrates the strength of robust architectures and data augmentation for improving diagnostic performance.

[9] Fahad Ahmed et al. employed the VGG-16 architecture along with explainable artificial intelligence techniques for brain tumor identification and prediction. Their model achieved 97.33% accuracy on a brain MRI dataset, emphasizing the value of transparency in AI models for medical imaging and ensuring interpretable, trustworthy results.

[10] Sulejman Karamehic and Samed Jukic focused on brain tumor detection using DenseNet and MobileNet architectures, with the support of the Python Imaging Library. DenseNet, in particular, showed superior performance, highlighting the importance of efficient and lightweight models for deployment in resource-limited environments.

[11] Shenbagarajan Anantharajan et al. worked on brain tumor detection from MRI images, combining deep learning with traditional machine learning techniques. Their model, trained on a dataset of 255 T1-mode MRI images, demonstrated the advantage of blending classical and modern approaches, especially for small-scale datasets.

[12] Sonia Ben Brahmin, Samia Dardouri, and Ridha Bouallegue utilized a deep CNN model for brain tumor detection, achieving an accuracy of 96% on a dataset of 3,000 MRI images. This research highlights the importance of diverse and comprehensive datasets for enhancing the generalization ability of diagnostic models.

[13] Aamir Muhammad et al. proposed an optimized CNN-based approach for brain tumor detection and classification, achieving a 97% accuracy across three Kaggle datasets. Their work underlines the importance of fine-tuning architectures to handle varying data sources effectively for better performance.

[14] Renhao Liu et al. focused on feature-based classification of brain tumor MRI images using deep convolutional neural networks (DCNN). Their model achieved an accuracy of 95.40%, stressing the importance of capturing detailed features in medical image analysis to ensure high accuracy.

[15] Yohan Jun et al. employed a 3D CNN model for detecting metastatic brain tumors, utilizing 3D black blood imaging. Their model achieved an accuracy of

97.08% on a clinical dataset, demonstrating the value of analyzing volumetric data for identifying complex tumor structures.

[16] J. Amin et al. developed a computer-assisted approach using brain MRI for distinguishing between malignant and benign lesions. Using an SVM classifier, they achieved 98% accuracy across multiple datasets, showcasing the flexibility of machine learning techniques in adapting to different medical datasets for accurate diagnoses.

[17] Gull S. and Akbar S. implemented a binary classification model for brain tumors using DenseNet and DarkNet-based CNN architectures, achieving an accuracy of 96.52% on the BRATS 2018 dataset. Their work demonstrates the potential of transfer learning to optimize performance on medical image datasets.

[18] Praveen Kumar Ramtekkar et al. proposed a CNN-based approach for brain tumor detection, achieving an accuracy of 98.9% on MRI images. This study emphasizes the benefits of end-to-end learning frameworks in streamlining medical diagnostic workflows.

[19] R. Rajasree et al. worked on brain tumor classification using deep learning models, including U-Net and CNN architectures. Their model achieved 96.36% accuracy on the BRATS 2015 dataset, showcasing the utility of segmentation networks in improving classification accuracy by isolating tumor regions effectively.

[20] S. Deepak et al. explored hybrid deep learning models for feature extraction, combining a CNN and SVM. Their approach, applied to the Figshare and brain MRI datasets, achieved 97.10% accuracy, emphasizing the complementary strengths of deep learning and traditional machine learning models for working with feature-rich data.

[21] Alexander S. Lundervold et al. used supervised learning techniques throughout the MRI processing pipeline, including acquisition, segmentation, and illness prediction. They achieved 95.20% accuracy by combining quantitative susceptibility mapping with CNNs, demonstrating how advanced imaging techniques can enhance deep learning performance.

[22] Arshia Rehman et al. developed a three-layer CNN model using transfer learning with VGG-16 for brain tumor detection. They achieved 98.69% accuracy on MRI brain slices from 233 patients, showing the effectiveness of transfer learning in reducing training time while maintaining high model performance.

[23] Justin S. Paul et al. applied various deep learning and machine learning techniques for brain tumor detection using CNNs. Their model achieved an accuracy of 91.40% on a tumor dataset, highlighting the challenges of dataset imbalance and the need for pre-processing techniques to improve model robustness.

[24] Md Abdullah Al Nasim et al. worked on tumor detection using a range of algorithms, including SVM, KNN, MLP, and others. SVM achieved the highest accuracy of 92.42% on a real-time dataset, demonstrating the reliability of SVM for multi-class classification tasks.

[25] Dewage et al. proposed custom CNN architectures alongside ResNet V2, DenseNet201, and VGG-16 for brain tumor detection. Their custom CNN model achieved an accuracy of 94.51% on a brain tumor MRI dataset, emphasizing the importance of model customization for improving efficiency and performance.

[26] Eze Benson et al. used CNNs to classify brain tumor features, achieving 92.00% accuracy on the BRATS 2018 dataset. Their research highlighted the significance of integrating feature engineering with deep learning for analyzing complex datasets.

[27] Fabian Isensee et al. applied a DCNN model with a U-Net architecture for brain tumor detection, achieving an accuracy of 90.10% on the BRATS challenge dataset. Their work showcased the ability of U-Net's encoder-decoder structure to handle both segmentation and classification tasks effectively.

[28] Takahashi et al. compares ViTs and CNNs in medical image analysis, evaluating their effectiveness in terms of robustness, efficiency, scalability, and accuracy. Analysis of 36 studies reveals that ViTs, especially with pre-training, often outperform CNNs in handling complex datasets. The findings aim to assist researchers in choosing suitable models for medical imaging tasks based on recent advancements.

## 3. PROPOSED SYSTEM

Journal of Theoretical and Applied Information Technology
15th February 2025. Vol.103. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

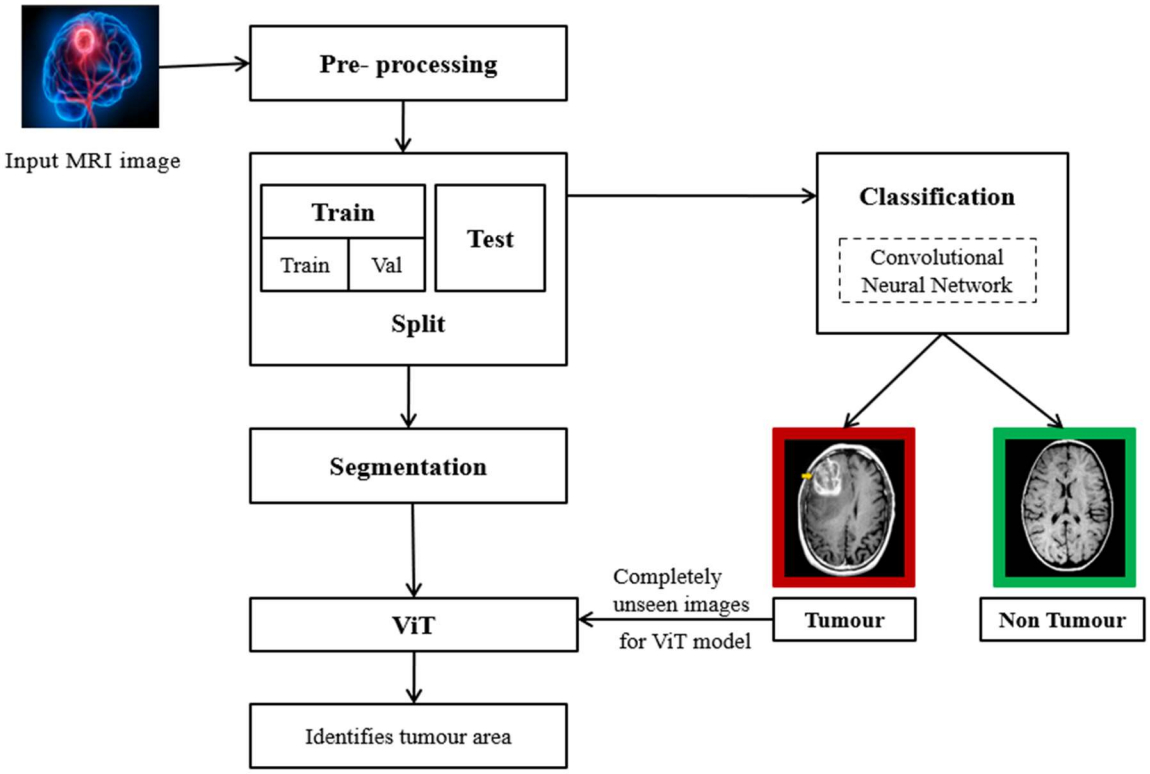


*Figure 1: Proposed Architecture for Brain Tumor Detection and Localization*

The proposed architecture for brain tumor detection uses a pipelined approach, integrating two distinct yet complementary sections for enhanced accuracy and precision. In the first section, a CNN-based classification model is utilized to accurately differentiate between tumor and non-tumor images, ensuring reliable identification. In parallel, the second section uses a ViT model to detect and locate the affected tumor area within the MRI images. The proposed framework in Figure 1. illustrates a step-by-step procedure outlining the various stages involved in tumor detection.

### 3.1 About Dataset

To train the models, the Br35H dataset obtained from the Kaggle repository has been used. There are two main sections in the dataset: one contains yes and no folders, where yes represents the tumor-contained MRI images, whereas no contains non-tumor images. Each of the folders contains 1500 images. The second section consists of train, test, and val folders with 500, 100, and 201 MRI images, respectively. Additionally, it includes an annotations_all.json file, which provides essential annotations for the images, including segmentation masks that highlight the tumor regions in the MRI images and labels that indicate whether an image contains a tumor or not, along with any additional segmentation information for training, validation, and testing purposes.

### 3.2 Pre-Processing MRI Images

The preprocessing step varies for each pipeline. For the ECNN model, preprocessing begins with resizing the images to 224×224 pixels.

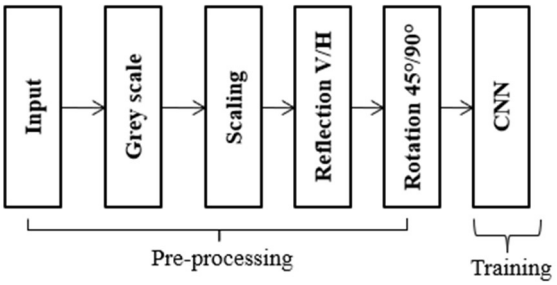


*Figure 2: Preprocessing for ECNN model*

Next, the input images are converted from RGB format to Grayscale to reduce the image complexity and number of channels in order to improve computational efficiency. Following this, Scaling is performed, which normalizes the pixel intensity values to a specific range by dividing the pixel values by 225, ensuring they fall within a specific range.

Later, data augmentation techniques such as Reflection, which creates the mirror versions of the images along vertical or horizontal axes, and Rotation, which rotates the images by specified angles, such as 45 or 90, are performed. Figure 2. explains the preprocessing steps of ECNN model. After pre-processing, the processed images are fed into the ECNN for training.

Similar to the ECNN model, the ViT model's preprocessing step begins with resizing all input images to a target size of 224×224 pixels to ensure uniform input dimensions for the model [29]. Next, the images undergo denoising to remove the unwanted noise and improve the image quality. Binary masks are then generated from annotations, where pixels corresponding to regions of interest, such as tumors, are set to 1, and the rest of the pixels are set to 0. The images with their respective masks are resized and normalized by dividing pixel intensity values by 255, scaling them to a range of [0, 1], to align with model requirements and enhance computational efficiency. Annotation files in JSON format are loaded for training, testing, and validation datasets to ensure alignment between images, masks, and ground truth. Each resized image is further divided into 16×16 non-overlapping patches, resulting in 196 patches, which serve as input tokens for the Vision Transformer (ViT). This patching process enables the ViT to learn both local and global patterns, facilitating accurate segmentation by highlighting critical tumor regions and understanding relationships between distant areas within the image.

### 3.3 Classification and Tumor Identification

The dataset is divided into training, validation, and test subsets to ensure reliable model evaluation. For the ECNN model, the balanced dataset, consisting of an equal number of tumor and non-tumor images, is utilized. A total of 3000 images are allocated for training and testing, with various split ratios explored to identify the optimal performance. The best results are achieved with an 84:16 train-test split. Furthermore, the training set is subdivided, with 84% used for model training and 16% reserved for validation, facilitating cross-validation. Finally, the ECNN model is evaluated on unseen test data to classify images accurately as either tumor or non-tumor.

For the Vision Transformer (ViT) model, another section of the dataset, pre-divided into training, validation, and test subsets, is utilized. The ViT model and U-Net are trained for tumor segmentation using the ground truth masks provided in the dataset. Tumor segmentation involves identifying and localizing tumor regions within MRI images. Semantic segmentation is used to classify each pixel in the image as either part of the tumor or background. The U-Net model, a widely recognized architecture in medical imaging, processes 224x224 input images through a symmetric U-shaped network. It consists of an encoder (contracting path) to capture features and context and a decoder (expanding path) to reconstruct segmentation maps. It outputs a segmentation mask of size 224x224, highlighting the tumor regions. Both the ViT and U-Net models are trained for 10 epochs with a batch size of 16. Post-training, validation accuracy across epochs is compared, and the ViT model outperforms U-Net in segmentation accuracy. As a result, the ViT model is employed to segment tumor regions in MRI images classified as tumors by the ECNN model, ensuring precise tumor location.

## 4. DEEP LEARNING MODELS FOR TUMOR CLASSIFICATION AND LOCALIZATION

In this article, deep learning models are used for detecting and locating the tumors in brain MRI images, which address both classification and segmentation tasks. MRI images are classified into tumor and non-tumor categories with great accuracy using Enhanced CNNs. For the second task of locating tumors, advanced architectures like ViT and UNet are compared to identify the best model for tumor localization. These models are designed to assist radiologists in the precise detection and identification of tumor regions in their early stages, enhancing diagnostic accuracy and facilitating more informed and effective clinical decision-making.

### 4.1 Enhanced CNN(ECNN)

The architecture of the ECNN model is illustrated in Figure 3, showcasing all the layers along with their respective details, including layer types, configurations, and parameters.

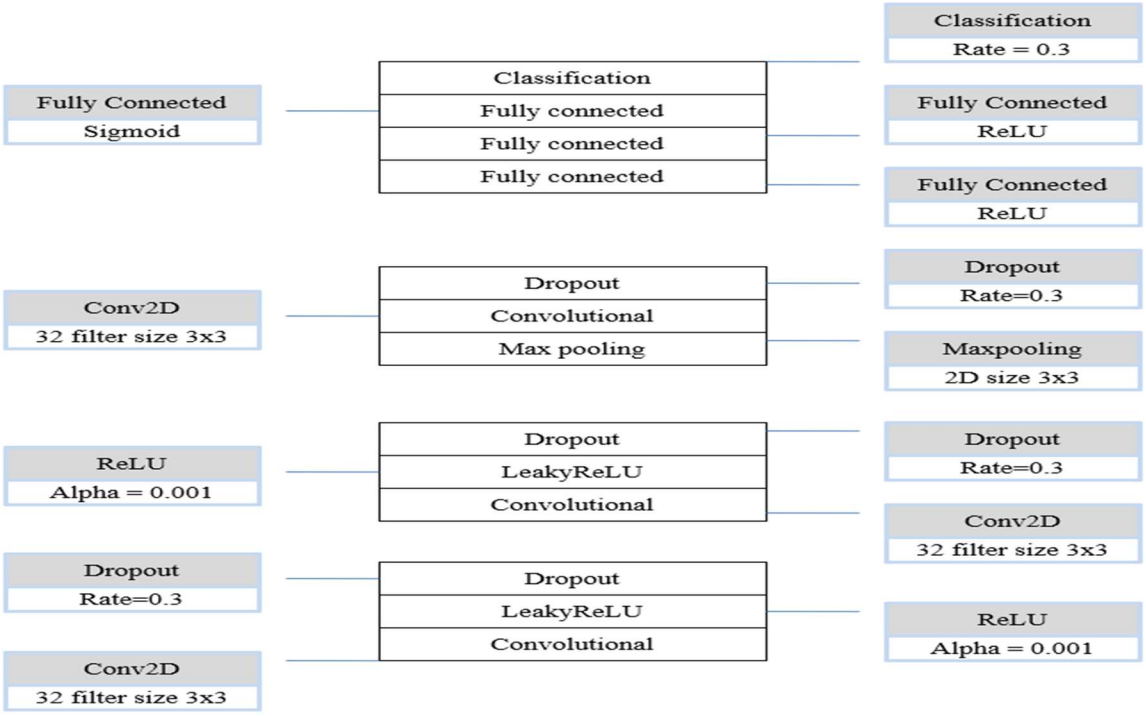


*Figure 3: Architecture of ECNN model*

The network begins with an input layer that accepts MRI images, which are typically resized and normalized to ensure consistent processing. The network's first layer is a Convolutional (Conv2D) layer with 32 filters, each measuring 3x3 in size. The purpose of this layer is to extract essential features, such as edges or simple textures, from the input image by applying convolutional operations. The convolution operation is mathematically expressed as shown in equation (1):

$$Z = \sum_{[i=1]}^{[n]} \sum_{[j=1]}^{[m]} I[i,j] \cdot K[i,j] + b \quad (1)$$

where $I$ represents the input image, $K$ is the filter kernel, and $b$ is the bias term. This process outputs feature maps, highlighting regions of interest. Following the convolutional operation, the output is passed through a LeakyReLU activation function, defined as shown in equation (2):

$$f(x) = \{ x \; if \; x > 0 \quad \alpha x \; if \; x \leq 0 \quad (2)$$

where α is a small positive constant (commonly set to 0.001). LeakyReLU allows a small gradient for negative inputs, preventing the "dying neuron" problem and improving model learning during training. To address overfitting, a dropout layer is incorporated, randomly deactivating 30% of the neurons during training. This stochastic regularization enhances the model's ability to generalize to unseen data. Following this, the output is passed through a 2D max-pooling layer, which reduces spatial dimensions by retaining the maximum values from 3x3 regions. Max pooling serves to reduce dimensionality, improving computational efficiency while capturing dominant features and ensuring translation invariance.

The second convolutional block begins with another convolutional layer, similar to the first, using 32 filters of size 3x3 to extract more complex features and patterns, such as shapes and textures essential for tumor detection. This convolution operates as previously described. The LeakyReLU activation function is used to introduce non-linearity, allowing the model to capture complex relationships in the data. To minimize overfitting, a dropout layer with a rate of 30% is incorporated [30]. Following this, a max-pooling layer is applied to downsample the feature maps while retaining the most critical patterns. The structured stacking of convolutional, activation, dropout, and pooling layers allows the network to learn both low-level and high-level features crucial for accurate tumor classification.

Once feature extraction is complete, the network moves to the fully connected layers, where the extracted features are transformed into a 1D vector. This vector is processed through several fully connected layers, enabling high-level reasoning and decision-making. The purpose of these layers is to integrate all the extracted features and map them to a final classification output. Each fully connected layer employs a ReLU activation function to add non-linearity. This ensures faster convergence during training and reduces the likelihood of vanishing gradients. The final output from the last fully connected layer is processed through a sigmoid activation function, producing a probability score between 0 and 1. This score represents the likelihood of the input image being classified as either "tumor" or "non-tumor".

### 4.2 U-Net Model for Tumor Localization

Advanced deep learning models, such as U-Net and Vision Transformer (ViT), are selected after a detailed comparison of related studies. U-Net, a convolutional neural network (CNN) architecture, is tailored for biomedical image segmentation and consists of two primary parts: an encoder and a decoder. The encoder focuses on extracting features and understanding the context of input images, while the decoder reconstructs the spatial information to produce segmented outputs. The architecture begins with an input layer capable of handling 2D or 3D images, where pixel intensity values serve as input features. The encoder utilizes downsampling to extract key features through several convolutional blocks, each comprising two 3x3 convolutional layers activated by ReLU, followed by a 2x2 max-pooling layer with a stride of 2 for spatial dimensionality reduction. To improve the representation of high-level features, the number of feature channels doubles at every stage of the encoder.

The bottleneck phase, positioned between the encoder and decoder, includes two convolutional layers with ReLU activation and batch normalization to enhance feature extraction. The decoder mirrors the encoder, utilizing transposed convolutions to restore spatial dimensions. It also incorporates skip connections, which merge features from corresponding encoder layers to preserve spatial context. Each decoding block contains a transposed convolution layer and two 3x3 convolutional layers with ReLU activation. In the final stage, a 1x1 convolution layer in the output layer reduces the feature maps to align with the number of target classes, generating a segmentation mask where each pixel is assigned a class label. The model's effectiveness is measured using the accuracy metric and is compared with the performance of the ViT model to evaluate its efficiency.

### 4.3 ViT Model for Tumor Localization

Vision transformers function similarly to traditional transformers, utilizing the self-attention mechanism to identify the relevance of various elements within a sequence for accurate predictions, yielding excellent performance in sequence-based tasks. As shown in Figure 4, the ViT workflow starts by segmenting an image into smaller patches of fixed size, where each patch corresponds to a specific region of the image. The pixel values from each patch are flattened into a one-dimensional vector, allowing the model to handle the patches as sequential data. These vectors are then projected into a lower-dimensional space using trainable linear transformations, preserving critical features while reducing dimensionality.

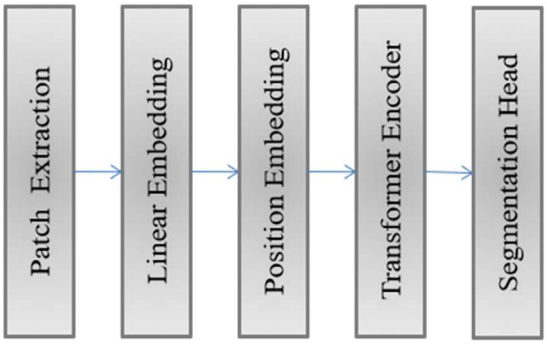


*Figure 4: Workflow of Vision Transformers*

Positional encodings are incorporated to preserve the spatial relationships between patches. These encodings help the model grasp the relative locations of various patches within the image. Mathematically, for a patch $x_i$, the embedding is computed as shown in equation (3):

$$z_i = W_p x_i + E_p \quad (3)$$

where Wp is the learnable projection matrix and Ep represents the positional embedding. The input to the encoder comprises the sequence of patch embeddings and positional embeddings.

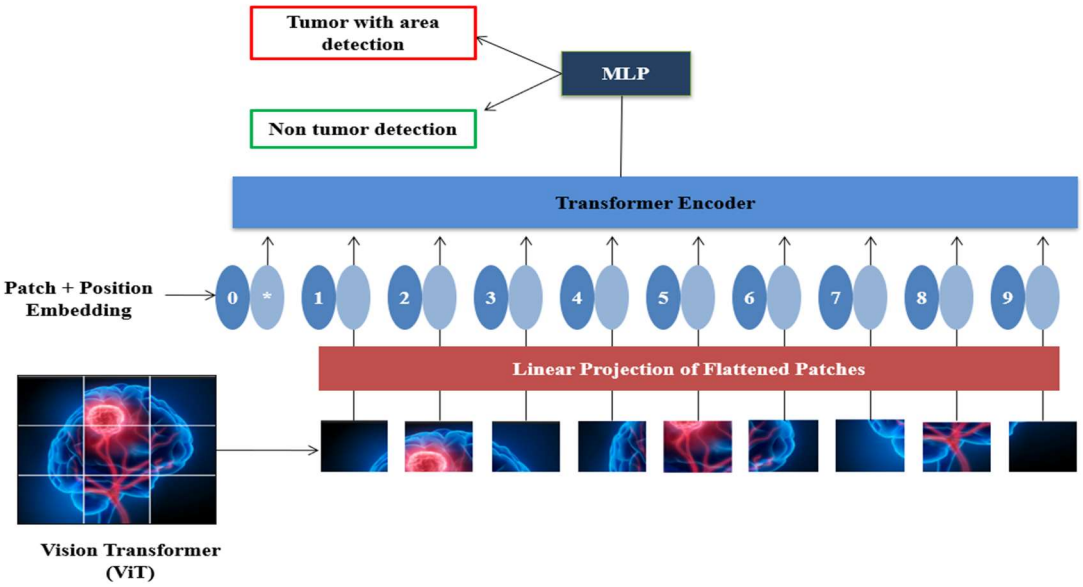


*Figure 5: ViT for Brain tumor detection*

The encoder consists of multi-head self-attention mechanisms (MHSA), multi-layer perceptron (MLP) blocks, and layer normalization (LN) applied before each block. MHSA computes relationships between patches by projecting the input into Query (Q), Key (K), and Value (V) matrices using learnable weights $W_Q$, $W_K$, and $W_V$. The attention scores are calculated as shown in equation (4):

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

where $d_k$ is the dimensionality of the keys. The Feed Forward Network (FFN) applies a two-layer fully connected network with non-linearity, given as shown in equation (5):

$$FFN(x) = W_2\,\sigma(W_1 x + b_1) + b_2 \quad (5)$$

where $W_1$, $W_2$, $b_1$, $b_2$ are learnable parameters, and σ is an activation function (e.g., ReLU or GELU). During training, an optimizer modifies the model's hyperparameters based on the loss calculated in each iteration. For image classification, a unique "classification token" is added to the beginning of the sequence of patch embeddings. The final state of this token, after passing through the transformer encoder, represents the overall image.

In the case of brain MRI images, images are converted into patches, which are provided as inputs to the patch embedding layer as shown in figure 5. The patch embedding layer converts patches into dense vectors and adds positional embeddings. A 16x16 patch is embedded into a 64-dimensional vector. The transformer block captures relationships between patches and processes attention outputs. Then upsampling transforms the output back to image dimensions (224 x 224) using transposed convolutions. The final output is a segmented image where each pixel is classified into specific categories, such as tumor or non-tumor regions. The upsampling step ensures that the spatial resolution of the segmented output matches the original input image dimensions, preserving details for accurate tumor localization.

## 5. RESULTS

The models are trained by using the Br35H dataset, which has different folders mapped to train different models. The ECNN models use the balanced dataset, which has 3000 tumor and non-tumor images. The graph in figure 6 illustrates the training and validation loss over 20 epochs. It is clear that both the validation and training losses exhibit a significant decline, indicating effective learning by the model. The training loss decreases steadily without overfitting, as reflected by the consistent behavior of the validation loss. By the 18th epoch, the validation loss stabilizes at a minimal value, signifying that the model has achieved optimal performance and is well generalized to unseen data.

Journal of Theoretical and Applied Information Technology
15th February 2025. Vol.103. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

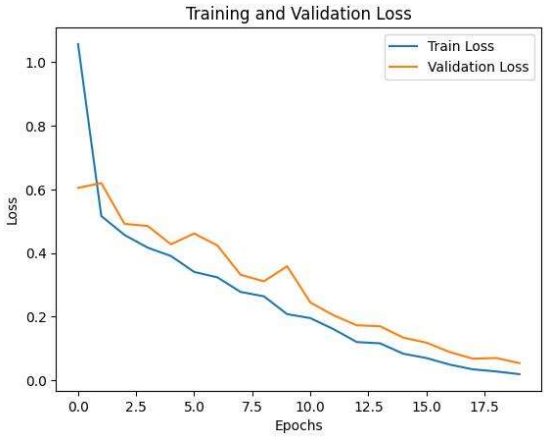


*Figure 6: ECNN Model Training and Validation Loss*

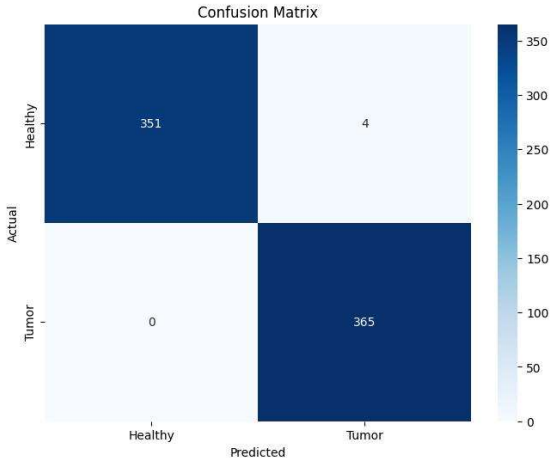


*Figure 7: ECNN Model Confusion Matrix*

Figure 7 is a confusion matrix that evaluates the model's classification performance. It shows that out of all healthy samples, 351 were correctly classified as "Healthy," and only 4 were misclassified. Similarly, for tumor samples, the model achieved perfect classification, identifying all 365 cases correctly. This results in a high classification accuracy, with near-zero false negatives, which is particularly crucial for tumor detection tasks. The table 1 shows the performance metrics achieved by the ECNN model.

*Table 1: ECNN Model Performance Metrics.*

| Metric | Score (%) |
|---|---|
| Accuracy | 99.4 |
| Precision | 98.9 |
| Recall | 1.00 |
| F1-score | 99.4 |

The ECNN model achieved a remarkable performance in classifying brain MRI images, with an accuracy of 99.4% and a precision of 98.9%.

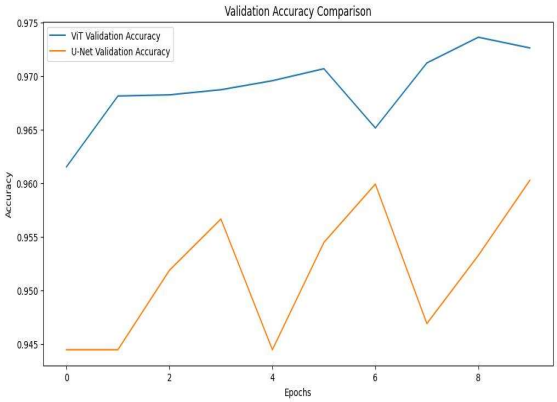


*Figure 8: Validation Accuracy of ViT vs U-Net*

The graph in figure 8 shows the comparison of the validation accuracy between the ViT and U-Net models and reveals that ViT consistently achieved higher accuracy across all epochs. Specifically, ViT maintained a validation accuracy ranging from approximately 0.965 to 0.975, outperforming U-Net, whose accuracy varied between 0.945 and 0.960. In terms of overall performance metrics, ViT demonstrated superior accuracy, recall, and F1-score, indicating its strength in identifying tumor regions in brain MRI images. However, U-Net exhibited slightly better precision than ViT, reflecting its ability to minimize false positives effectively. This comparison highlights the capability of ViT for segmentation tasks.

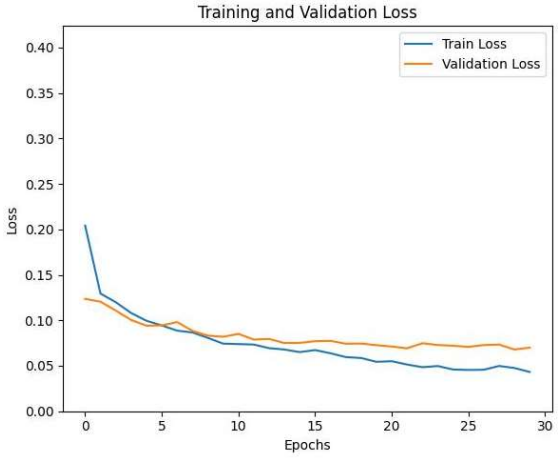


*Figure 9: ViT Model Training and Validation Loss*

Figure 9, illustrates the training and validation loss of ViT model over 30 epochs. These losses show good learning and convergence as they gradually decline. The close proximity of the two curves suggests minimal overfitting.

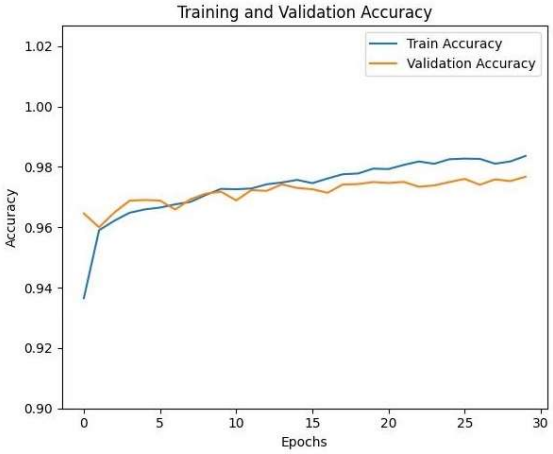


*Figure 10: ViT Model Training and Validation Accuracy*

The accuracy of the ViT model's training and validation over 30 epochs is shown in Figure 10, where both converge close to 98% and exhibit steady improvement. This showcases the ViT model's robust generalization to unseen validation data while achieving high accuracy, making it an excellent choice for segmentation tasks.


*Figure 11: ViT Model Training and Validation Accuracy*

Figure 11 illustrates the process of segmenting brain tumors from MRI scans. The first section shows the original MRI image, the second represents the actual tumor outline (ground truth mask), and the third highlights the model's predicted tumor region, emphasizing its precision.

## 6. CONCLUSION

This work explores the use of advanced deep learning models for detecting and locating brain tumors in MRI scans. The ECNN model demonstrated strong performance in accurately identifying the presence of tumors, proving effective for binary classification tasks. Additionally, a comparative evaluation of U-Net and Vision Transformer (ViT) models was performed for tumor region segmentation. While U-Net delivered reliable results for segmenting tumor areas, the ViT model surpassed it in accuracy, showcasing its superior ability to generalize across unseen data.

The ViT model's capability to leverage self-attention mechanisms and extract intricate features resulted in precise tumor localization, making it a valuable tool for clinical applications. Its outstanding performance demonstrates its ability to automate the segmentation of tumors in MRI images, providing radiologists with substantial assistance in diagnosis and treatment planning. This study also emphasizes the importance of advanced deep learning models like ViT in enhancing the accuracy and efficiency of medical imaging-based diagnostic systems.

## REFERENCES:

[1] Krishnapriya S., Karuna Y. "Pre-trained deep learning models for brain MRI image classification." Frontiers in Human Neuroscience, Vol. 17, April 20, 2023, pp. 1150120.

[2] Badža M. M., Barjaktarović M. Č. "Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network." Applied Sciences, Vol. 10, Issue 6, 2020, pp. 1999.

[3] Kumar S., Dabas C., Godara S. "Classification of brain MRI tumor images: A hybrid approach." Procedia Computer Science, Vol. 122, 2017, pp. 510–517.

[4] Abdallah Y. M. Y., Alqahtani T. "Research in medical imaging using image processing techniques." In: Medical Imaging—Principles and Applications, IntechOpen, 2019.

[5] Shen D., Wu G., Suk H.-I. "Deep learning in medical image analysis." Annual Review of Biomedical Engineering, Vol. 19, Issue 1, 2017, pp. 221–248.

[6] Litjens G., Kooi T., Bejnordi B. E., Setio A. A. A., Ciompi F., Ghafoorian M., et al. "A survey on deep learning in medical image analysis." Medical Image Analysis, Vol. 42, 2017, pp. 60–88.

[7] Saeedi S., Rezayi S., Keshavarz H., et al. "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques." BMC Medical Informatics and Decision Making, Vol. 23, 2023, Article 16.

[8] Srikanth B., Suryanarayana S. V. "Multi-class classification of brain tumor images using data augmentation with deep neural network." Materials Today: Proceedings, 2021.

[9] Ahmed F., Saleem M. A. M., Mushtaq U. F., Imram M. "Identification and prediction of brain tumor using VGG-16 empowered with explainable artificial intelligence."

[10] Karamehic S., Jukic S. "Brain tumour detection and classification using deep learning algorithm and Python imaging library."

[11] Anantharajan S., Gunasekaran S., Venkatesh R. "MRI brain tumor detection using deep learning and machine learning approaches." Measurement: Sensors, 2024. doi: 10.1016/j.measen.2024.101026.

[12] Ben Brahmin S., Dardouri S., Bouallegue R. "A novel approach for brain tumor classification using computational intelligence." Applied Computational Intelligence and Soft Computing, Vol. 2024, Issue 1, September 18, 2024, Article 7634426.

[13] Aamir M., Namoun A., Munir S., et al. "Brain tumor detection and classification using an optimized convolutional neural network." Diagnostics, Vol. 14, Issue 16, 2024, Article 1714.

[14] Liu R., et al. "Exploring deep features from brain tumor magnetic resonance images via transfer learning." International Joint Conference on Neural Networks (IJCNN), 2016, pp. 235–242.

[15] Jun Y., et al. "Deep-learned 3D black-blood imaging using automatic labelling technique and 3D convolutional neural networks for detecting metastatic brain tumors." Scientific Reports, Vol. 8, 2018, pp. 9450–9461.

[16] Amin J., et al. "A distinctive approach in brain tumor detection and classification using MRI." Pattern Recognition Letters, Vol. 139, 2017, pp. 1–10. doi: 10.1016/j.patrec.2017.10.036.

[17] Gull S., et al. "An interactive deep learning approach for brain tumor detection through 3D-magnetic resonance images." International Conference on Frontiers of Information Technology, 2021, pp. 114–119.

[18] Ramtekkar P. K., et al. "Innovative brain tumor detection using optimized deep learning techniques." International Journal of System Assurance Engineering and Management, 2022.

[19] Rajasree R., et al. "Multiscale-based multimodal image classification of brain tumor using deep learning method." Neural Computing and Applications, Vol. 33, 2021, pp. 5543–5553.

[20] Deepak S., Ameer P. M. "Brain tumor classification using deep CNN features via transfer learning." Computers in Biology and Medicine, 2019, pp. 1–7. doi: 10.1016/j.compbiomed.2019.103345.

[21] Lundervold A. S., Lundervold A. "An overview of deep learning in medical imaging focusing on MRI." Zeitschrift für Medizinische Physik, Vol. 29, Issue 2, 2019, pp. 102–127.

[22] Rehman A., et al. "A deep learning-based framework for automatic brain tumors classification using transfer learning." Circuits Systems and Signal Processing, Vol. 39, 2020, pp.757–775. doi: 10.1007/s00034-019-01246-3.

[23] Paul J., et al. "Deep learning for brain tumor classification." Biomedical Applications in Molecular Imaging, Vol. 10137, 2017, pp. 1013710–1013726. doi: 10.1117/12.2254195.

[24] Nasim M. A. A., Shah F. M., Hossain T., et al. "Brain tumor detection using convolutional neural network." June 2019.

[25] Dewage K. A. K. W., Hasan R., Rehman B., Mahmood S. "Enhancing brain tumor detection through custom convolutional neural networks and interpretability-driven analysis." Information, Vol. 15, Issue 10, 2024, Article 653. doi: 10.3390/info15100653.

[26] Benson E., et al. "Deep hourglass for brain tumor segmentation." International MICCAI Brainlesion Workshop, Vol. 10, Issue 2, 2018, pp. 419–428. doi: 10.3390/brainsci10020118.

[27] Isensee F., et al. "Brain tumor segmentation using large receptive field deep convolutional neural networks." Medizinische Informatik und Statistik, 2017, pp. 86–91.

[28] Takahashi, Satoshi & Sakaguchi, Yusuke & Kouno, Nobuji & Takasawa, Ken & Ishizu, Kenichi & Akagi, Yu & Aoyama, Rina & Teraya, Naoki & Shinkai, Norio & Machino, Hidenori & Kobayashi, Kazuma & Asada, Ken & Komatsu, Masaaki & Kaneko, Syuzo & Sugiyama, Masashi & Hamamoto, Ryuji. Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review. Journal of Medical Systems,2024, 48:84.

[29] Arati Rath, Bhabani Shankar Prasad Mishra, Dilip Kumar Bagal, "ResNet50-based Deep Learning model for accurate brain tumor detection in MRI scans", Next Research, Volume 2, Issue 1, 2025, pp. 100-104,

[30] Dangety Sowjanya, Dr. Kunjam Nageswara Rao, Dr.G. Sita Ratnam, K. Venkateswara Rao, "Covid-19 Prediction Using Transfer-Learning On Rt-Pcr Confirmed Cxr-Images", Indian Journal Of Computer Science And Engineering, 2022, Vol. 13, No.2, pp. 379-387.