iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

# BRAIN TUMOR DETECTION USING 3D U-NET SEGMENTATION AND EXPLAINABLE AI: A REVIEW TO IMPROVE ACCURACY AND TRANSPARENCY IN CLINICAL DIAGNOSIS

**Mr. Tejeshwarsingh H Kushwah**
PG Students, CSMSS College of Engineering,
Dr. Babasaheb Ambedkar Technological University (DBATU)-Lonere, Aurangabad, Maharashtra.

**Prof. Yogesh R Tayade**
Associate Professor, CSMSS College of Engineering,
Dr. Babasaheb Ambedkar Technological University (DBATU)-Lonere, Aurangabad, Maharashtra

**Prof. Nikhil M Sapate**
Assistant Professor, CSMSS College of Engineering,
Dr. Babasaheb Ambedkar Technological University (DBATU)-Lonere, Aurangabad, Maharashtra.

**Dr. Shrinivas R Zanwar**
Head of Department, CSMSS College of Engineering,
Dr. Babasaheb Ambedkar Technological University (DBATU)-Lonere, Aurangabad, Maharashtra.

**ABSTRACT**
Brain tumors represent an eminent domain of life threatening neurologic disorders where early diagnosis and correct diagnosis plays a central role in improving a survival outcome and in guiding a therapeutic treatment strategy. Due to its capacity to provide high-quality visualization of cerebral parenchyma and avoiding ionising exposure to the patients, Magnetic Resonance Imaging has become the diagnostic modality of choice in intracranial neoplasma[1]. However, traditional radiological interpretation of magnetic resonance images is often labour intensive, prone to intra- and inter-observer variability and prone to misdiagnosis, especially when larger volumes of data are involved.
More recent developments in the sphere of the artificial intelligence and deep-learning have provided a new wave of use of AI, which allows the fully automated process of the identification of a tumour and its delineation in MRI images [2]. Among the many options available in the deep-learning solutions, convolutional neural networks have demonstrated better performance in a wide range of radiological issues. Of special importance is the U-Net framework that has become a de facto standard in biomedical segmentation tasks.[3].
Canonical U-Net U-Net design takes advantage of the encoder-decoder architecture with skip connections, thus allowing to extract contextual semantics and at the same time learn fine-grained spatial detail. Although the initial definition was in two dimensions of imagery, more recent studies have superimposed a shift to volumetric representations, resulting in 3-D U-Net based systems[4]. These volumetric models embed entire-volume MRI information as opposed to the use of separate
axial slices, thus, representing inter-slice dependencies and significantly increasing segmentation precision of tumors.
**Keywords:**
Brain Tumor Detection, Medical Image Segmentation, 3D U-Net, Explainable Artificial Intelligence (XAI), Magnetic Resonance Imaging (MRI), Deep Learning, Medical Image Analysis, Grad-CAM, SHAP, Healthcare Artificial Intelligence, Computer-Aided Diagnosis, Multimodal MRI, Brain Tumor Segmentation, Clinical Decision Support Systems, AI Interpretability.

## INTRODUCTION

Brain tumors are the pathological overgrowth of cells located in the cerebral parenchyma or its surrounding tissues, which has a significant effect on the cerebral functions and health conditions overall of the subjects. Epidemiological statistics have shown that the neoplasms contribute a significant portion of cancers-related

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

deaths, across the world. Timely and accurate diagnosis of brain tumors is essential towards the formulation of the best therapeutic strategies, including surgical resection, radiotherapeutic, as well as chemotherapeutic institutions. However, the diagnostics remains a complex process, which can be explained by the heterogeneity of tumours, differences in their sizes and morphology, the complexity of the cerebral anatomy.

The Magnetic Resonance imaging (MRI) has been the modality of interest in the imaging of brain tumours, due to its better definition of soft tissue structure. MRI provides a way of gaining various image contrasts such as T1-weighted, T2-weighted, contrast-enhanced T1 (T1c), and Fluid Attenuated Inversion Recovery (FLAIR) images. Every contrast highlights various histological characteristics and, therefore, leads to a more detailed visualisation of tumours. In reality, radiological evaluation is generally done manually, which enables the specialists to outline tumor pathology and deduce histopathological subtypes. Nevertheless, manual inspection of MRI datasets is tedious, time consuming and fraught with high inter-observer discordance, which may result in diagnostic discordance.

The emergence of artificial intelligence has provoked new opportunities of automation of medical image interpretation. The use of machine learning methods in health-related spheres has been practiced on a wide scale and covers such aspects as the prediction of various diseases, the analysis of images, and the customization of treatment methods.

Traditionally, many methods of machine learning with regard to brain tumour detection were based on hand-made descriptions, including texture measures, histogram statistics, or geometric shape. These human-made features were also used as inputs to support vectors machines, random forests and k nearest neighbors classifiers. Inasmuch as these methods were partly successful, they often did not capture the wavy space distribution of MRI data. Deep learning protocols have significantly improved the effectiveness of image reading mechanisms, mainly by automating the extraction of features in unprocessed pixel data.

The ability to learn hierarchies in complex pictorials has made Convolutional Neural Networks (CNNs) the most popular paradigm of visually oriented tasks. Brain tumour detection CNN-based architectures identify pattern specific to tumour by automatically identifying this pattern in MR images bypassing manual feature engineering. One of the most effective architectures in the field of biomedical image segmentation is the U-Net architecture. U-Net was originally conceived based on the
division of the cell object in microscopic images, and the architecture is characterized by a symmetric encoder-decoder layout along with skip connections to combine contextual with spatial details.

The encoder phase retrieves high-level representations by a series of convolutional and pooling aims whereas the decoder recreates segmentation maps with the help of upsampling processes. Skip connections allow to combine small-scale spatial features with large context features, and thus, increase the accuracy of the final segmentation.

## RELATED WORK

Detection of brain tumor using medical imaging has emerged as one of the principal scientific investigations due to the growing rate of neurological conditions and the growing possibilities of the medical imaging technology. The Magnetic Resonance Imaging (MRI) technique, especially, provides an extremely high level of soft-tissue contrast, which allows clinicians to outline the pathological areas with an unprecedented level of granularity. the manual analysis of MRI images is tough and prone to get errors from observer ; therefore, this area is converted and shifted to fully automated, data-driven methods involving both machine learning and deep learning to improve accuracy in diagnostic tests and efficiency in the work process [1], [2].

Initial work in automated tumor detection consisted of traditional image-processing algorithms including thresholding and region growing and clustering and edge detection algorithms intended to single out tumor regions by means of intensity contrasts and geometric features in MRI images. These techniques produced early evidence of concept but were often weak in nature and could not be capable of accommodating intricate morphology that is typical of malignant lesions. Furthermore, the handcrafted dependence on the domain knowledge usually hampered generalizing to the various imaging cohorts that appeared in practice [3]. In order to overcome these drawbacks, researchers have turned to the use of supervised machine-learning classifiers, including Support Vector Machines (SVM), Decision Trees, Random Forests and k-nearest Neighbour (KNN) models, which are trained on manually-constructed descriptors, like texture, shape and intensity statistics obtained using MRI volumes. Experiments on comparative studies have shown that these algorithms were capable of outperforming straight image-processing baselines. However, they were limited in their ability to automatically acquire the complex, high-dimensional structures inherent to pathological imaging data because they relied on manual feature engineering [4], [5].

With the emergence of deep learning, medical analysis of images has been radically changed. Convolutional

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Neural Networks (CNNs) are recently at the state-of-the-art on a continuum of computer-vision problems: the key ones being object detection, image classification, and segmentation: through the automatic finding of hierarchical feature maps based on raw pixel values. This representational force has prompted the intensive use of CNN-based approaches to detecting and classifying brain tumors in publicly accessible MRI samples [6], [7]. The different CNN backbones such as VGGNet, AlexNet, ResNet, DenseNet, and EfficientNet have been used to distinguish between tumorous and non-tumorous voxel in MRI images. The use of residual networks (ResNet) in particular has attracted such interest because of their ability to run significantly deep architectures without the vanishing gradients. In practice an empirical evidence suggests that the classification accuracies of such models can reach up to over ninety percent of benchmark MRI corpora [8], [9].

Although CNNs are sufficiently useful in classification, clinical practice requires tumour boundaries to be localised accurately in three-dimensional imaging volumes. This has led to the creation of complex segmentation architecture that can mark the extent of tumour at the voxel level. The U-Net network which was presented as an innovative architecture of segmenting biomedical images is based on the
encoder-decoder architecture where the encoder progressively acquires feature representations at initially lower spatial resolutions (aggregation of features) and the decoder gradually learns to synthesize the finer spatial details to generate high-resolution segmentation maps [10]. One design capability that contributes to the success of U-Net is its skip-connection design, in which encoder layer maps bypass decoder layers with volumes of identical features. These shortcuts maintain fines grain spatial details that would otherwise be lost due to down-sampling and thus the model is able to faithfully recreate fine anatomical structures. Therefore, U-Net has emerged as an omnipresent backbone of biomedical segmentation, such as the delineation of brain tumors, organ localization and lesion localization, thanks to its beautiful tradeoff between depth and spatial fidelity [11]. Although it can be successfully applied to the 2D image segmentation, the naturally three-dimensional character of MRI data restricts the use of planes-wise U-Net by ignoring the context in the following cuts. To address the same deficiency, scholars have suggested the 3D U-Net architecture that builds on the original design by substituting the two-dimensional convolutions with the volumetric convolutions. This volumetric extension allows the model to consume and run whole MRI volumes, providing spatial constraints between more than two slices and hence providing more coherent and anatomically consistent segmentation solutions [12].

The 3D U-Net models have repeatedly demonstrated a higher performance in segmenting brain tumours than 2D based, especially because they have the ability to question the entire volumetric structure of tumours. With these models offering the ability to bring the network into contact with spatially adjacent slices, this artificially provides greater capability to tell tumours apart of adjoining neural tissue, which has been demonstrated through models based on the Brain Tumor Segmentation (BraTS) dataset where 3D U-Nets achieve high levels of Dice similarity indices and relative segmentation fidelity compared to traditional 2-D frameworks [13]. In addition to the classic 3D U -Net architecture, a collection of advanced designs has appeared, designed to drive the performance of segmentation at its limits. Residual U-Nets add shortcut correlations that allow extended back-propagation paths, thus boosting the depth of representations.

U-Nets have gating schemes that select salient parts as the foreground background is suppressed unlike Dense U-Nets where features are propagated using a dense inter-layer connectivity topology which actively promotes feature reuse and gradient concentration throughout the network [14], [15]. Another confounding factor of detection efficacy is the quality of datasets available. This has been enabled by the BraTS challenge, which has made available a thoroughly curated multimodal MRI corpus of acquisitions which contains T1, T1 -contrast (T1c), T2, FLAIR, along with expert-labeled tumoral localization. These benchmarked resources provide the researchers with a reproducible platform to assess segmentation algorithms based on such metrics as Dice coefficient, sensitivity, specificity, and Intersection over Union (IoU) [16].

Although deep learning models have significant advances in the accuracy of prediction, the models remain opaque and therefore do not support clinical adoption. Deep neural networks are black-box and therefore it is hard to explain what is going on inside the computer as it is being predicted. Medical decision individuals on the other hand require systems that do not only make a prediction but also provide explanations, which has driven the development of Explainable Artificial Intelligence (XAI) systems that attempt to make deep learning processes transparent [17]. GradientWeighted Class Activation Mapping (Grad-CAM) has become popular as one of the XAI tools due to its capability of generating heatmaps onto the input images to highlight the parts that produce the most impact on a classification task. Grad-CAM may be used to detect tumor localisation in MRI images in the context of tumor detection, providing clinicians with the option of determining whether the model is directing attention where it should to clinical anatomy [18]. Another popular method is the Local Interpretable Model-Agnostic Explanations (LIME), an approximation method that calculates the output of a complex model

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

locally by a model that is simple and interpretable. LIME can be used to examine the impact of discrete image patches on the final prediction by perturbing the input data and comparing the resulting change in predictions, which allows learning increasingly about the decision hierarchy of a model [19].

Based on the cooperative game theory, Shapley Additive Explanations (SHAP) provides a theoretically sound, coherent way of allocating the contribution to prediction based on the contribution of individual features. The SHAP scores can be applied to the medical imaging, where they associate the MRI voxels that contribute to the tumour detection the most, which supports the reliability and interpretability of the system [20]. Recent studies are moving towards the direction of combining these explainability methods with more advanced architectures like the 3D U-Net. Researchers can binarily identify tumour regions and understand the rationale behind each prediction with XAI modules directly integrated into segmentation pipelines, and this dual function is the most critical step towards inspiring clinical confidence in AI-enhanced diagnostics.

However, the discipline addresses long-standing predicaments. One of the bottlenecks is the lack of highly annotated medical datasets: the labeling of high quality requires the participation of skilled radiologists, which makes the process data collection resource-consuming and time-consuming. This results in most gain access to deep learning models having been trained on small sample sizes and are thus more susceptible to overfitting and a lack of generalisability. The second barrier is the huge computational costs of volumetric network training, which require large memory footprints, and computing capabilities, which could inhibit their use in clinics with time constraints. With light-weight architecture and model compression algorithms, the community is actively developing and investigating alternative ways of reducing these limitations. In short the literature shows clearly that deep learning has significantly improved the speed of brain tumour detection and segmentation, acceleration in the accuracy of these processes. Headers to U-Net, especially when applied in three dimensions, have recorded impressive achievements in medical imaging.

Simultaneously, the interpretability gap has been mitigated by XAI tools, including Grad-CAM, LIME, and SHAP, which provide visual stories that can be used to explain algorithmic decision-making to clinicians. However, data (lack) and the computational overhead (time) and transparency considerations remain as high-priority research challenges. An effective remedy to the mentioned gaps by a conciliating approach of 3D U-Net segmentation and explainable AI is the proposed project, and the long-term objective is to present a suitable, precise, and clinically deployable brain tumour detection system.

A detailed attention was given to academic studies that utilized 3D U-Net models in delineation of brain tumours. The models were critically assessed with regard to their ability to consume volumetric MRI images, as well as define spatial dependencies among serial slices of images. The accompanying advantages and shortcomings of these architectures were analyzed at length.[39]. In addition to the segmentation paradigm, explainable artificial intelligence methodologies were also covered in the review.

Methods like Grad-CAM and SHAP were evaluated in terms of the effectiveness in improving the transparency of these models and providing readable details of AI-driven insights into the predictions generated by them [40]. The usefulness of the methods in the medical imaging field was proven by the exemplary case studies based on the available literature. After that, a theoretical framework is described that integrates the 3D U-Net segmentation stream with explainable AI practices. The

proposed model aims to address the failure flaw of existing methodologies through integrating accurate tumour localisation with clear model justifications.[42]. Such a combination promises to increase the reliability and the clinical usability of AI based brain tumour sensing systems.

U-Net with its variations is one of the best deep learning structures used in segmentation tasks with consistent good results. They have reported Dice scores exceeding 90 0 -percent with Tumour segmentation using U-Net-based models. Nonetheless, recent models like attention-based networks and transformer-based models have shown additional gains in the precision of segmentation. The attention mechanisms allow the model to focus on the important regions of the tumours and disregard the background noises. Explainable AI methods are important concerning the assessment of the credibility of these models. Visualization software like Grad-CAM has the capability of elucidating the predicates of the model by showing the areas of MRI images that the model relies on. According to the comparative analysis of the latest works, it seems that models that combine 3D segmentation architectures and explainability frameworks are the most appropriate ones in terms of accuracy and explainability. This has come with new challenges with the growing complexity of deep learning models.

Among the most serious challenges, there is the fact that deep learning predictions are not interpretable. Medical practitioners need clear definitions of diagnostic actions, especially when diagnostic actions are related to life-threatening situations like brain tumours. Clinicians do not trust black-box AI systems which generate predictions without explanations and restrict their use in healthcare settings. Explainable Artificial Intelligence

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

(XAI) is a promising method of resolving such a problem. The XAI methods seek to offer an understanding of the internal decision-making mechanism of the machine learning models. Efforts like Grad-CAM produce visual heatmaps indicating areas within an image that contribute when making predictions with the model against the model. Likewise,SHAP values determine the importance of individual features to the output of the model. Such methods enable the clinicians to know the manner in which AI models perceive the medical images and can also test the accuracy of predictions. The high complexity of deep learning models is another limitation of brain tumour detection research because their computation cost is prohibitive.

The computationally resources needed to train large neural networks and data will be huge and are not necessarily accessible in most healthcare facilities. Scientists are consequently examining less

efficient structures and optimisation strategies to curtail the number of calculations. Moreover, the scarcity of various medical imaging data presents a major challenge towards coming up with more powerful AI models. The BraTS dataset is extremely important in numerous studies, yet it was found not to be as diverse a sample of clinical imaging data as it could be. Future studies should be dedicated to the development of bigger and more varied datasets to enhance generalisation of models. The aim of this review paper is to examine the latest advancements in detection of brain tumours with the help of 3 -D U-Net segmentation models and explainable AI methods. This paper will offer informative findings into future research pathways to create a more accurate, interpretable, and practical brain tumour detection device by analysing current literature and pinpointing major limitations. Once the first set of the research articles is collected, the first type of filtering was introduced to see the most relevant studies.

## OBJECTIVES

The papers that specifically targeted deep -learning based detection of brain tumours through MRI data were included in the analysis. The review did not include studies that employed alternative imaging methods or machine learning methods that were not related. The brain tumour detection by AI is examined in this paper using a summary of this topic and an overview of this research paper. The application of artificial intelligence in medical imaging has yielded one of the most important uses in brain tumour detection. Brain tumours are the un-normal growths of cells in the brain tissues that may have serious implication on the on the functionality of the brain and survival of patients. The identification of tumour foci and their precise classification at the earliest stage is very important in creating treatment programs and tracking the evolution of the disease.

The magnetic Resonance Imaging (MRI) has become common in the diagnosis of brain tumours due to the fact that it presents high-resolution images of brain tissues, and the various images that can be generated using this modality of imaging include T1, T2, T1-contrast (T1c), and Fluid-Attenuated Inversion Recovery (FLAIR). The first use of machine learning methods was in automated detection of brain tumours. Initial methods used handcrafted characteristics which were produced based on MRI data that included texture characteristics, intensity histograms, and shape characteristics. The inputs were then fed into classifiers like Support Vector Machines (SVM), Random Forests and KNearst Neighbor. Even though these approaches showed moderate success, they were constrained by the quality of manual designed features and the inability to measure complicated spatial patterns of MRI pictures. Deep learning has played a major role in medical image analysis, making it possible to extract automatic features on raw data of images.

The prolific architecture of CNNs training hierarchical visual patterns made them the architectural approach of choice in image-based tasks. CNN models do not need any features to be engineered manually since they are capable of automatically identifying tumour-related features in MRI. These models have realised considerable breakthrough in tumour location accuracy over the conventional machine learning methods. Nevertheless, early CNN designs were primarily aimed at classification of images and not at pixel-wise tumour region segmentation. Proper segmentation is required in the determination of tumour size, shape, and location; which are vital variables in the planning of treatment. As a result of this limitation, specialised segmentation architectures have been developed like U-Net.

### U-Net in Medical Image Segmentation.

U-Net architecture has been firstly introduced with the purpose of biomedical image segmentation, and it has made one of the most popular models to use medical images tasks. The architecture entails a symmetric encoder-decoder design that acquires contextual and spatial information of images. The encoder obtains features of hierarchy with convolution and pooling processes, and the decoder restores segmentation maps with up-sampling layers. The application of skip connections between encoder layers and the corresponding decoder layers is one of the most important inventions of the U-Net architecture. These links enable the model to

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

integrate semantic information at the high level with the spatial details of low level and thus enhance the accuracy of the segmentation.
Consequently, U-Net models have been used extensively to the segmentation of tumors, organs, and other body parts in medical images. It has been revealed that U-Net structures are capable of high-level accuracy of segmentation when used in brain tumour detection tasks. Indicatively, the studies have revealed that U-Net based models can reach a level of segmentation accuracy and a Dice score of beyond 90 per cent when utilized with MRI Tissue data like BraTS. In spite of these achievements,
conventional U-Net-based models operate upon images in two dimensions (slices) and not a complete volumetric scan of MRI. This drawback could lead to the blanking out of spatial information between slices by the model, which could lead to a lower level of segmentation. In order to overcome this problem, authors proposed 3D U-Net networks.

**3-D U-Net on Brain tumor Segmentation.**
Deep-learning models, which are based on 3D, have found their way into brain tumour segmentation. The 3D U-Net architecture is a modified version of the U-Net as it uses 3D convolution feature instead of 2D convolution feature and enables the model to learn volumetric features of MRI data and learn spatial relationship across neighboring slices. Services provided by 3-D U-Net models
The fact that full volumes of MRI can be analysed makes 3-D U-Net models more likely to reconcile tumour morphology and distances among brain tissues and therefore enhance the level of segmentation and also give more accurate tumur boundaries. A number of studies have established the efficiency of 3-D U-net models on the detection of brain tumours. Scholars have considered modified U-Net structures to the BraTS information set and demonstrated high segmentation results in various sub-areas of tumours that include the entire tumour, tumour core, and enhancing tumour. Not so old studies have also suggested better 3-D U-Net networks with attention models, residual networks, and multi-scale feature domain to improve results on segmentation. As an example, attention-based 3-D U-Net models have been found to the ability to attain high Dice scores and gain superior sensitivity by concentrating on applicant tumour areas and oppress irrelevant background details. The other technique incorporates simultaneous use of decoder architecture and attention-controlled skip connections to enhance efficiency and low computational needs of segmentation. These models have shown performance of competitive segmentation on BraTS data set with less training epoch and less computational resources. Despite the fact that 3D U-Net architectures achieve significant enhancement in the accuracy of segmentation, they also imply a new set of problems connected with the complexity of computations and the interpretability of models.

**Medical imaging Explainable Artificial Intelligence.**
Deep-learning-based systems can frequently work as black-box systems i.e. their decision-making mechanisms are challenging to explain. Such opiqueness may be an issue in medical practice due to the fact that clinicians should be aware of the rationale behind diagnostic decision-making. Explainable Artificial Intelligence (XAI) has become a significant field of study to enhance understanding of machine-learning models. The XAI methods contribute to the understanding of how models can make a prediction, and guide attention areas of the model, by focusing on meaningful features and illustrating attention areas in the model. Gradient-weighted Class Activation Mapping (Grad-CAM), SHAP, and LIME are all explainability methods that are normally employed in the detection of brain tumours. Grad -CAM creates heatmaps, which points out the parts of MRI images which affect the predictions of the model whereas SHAP values are feature-importance scores that indicate the contribution of each input changes the output of the model.
It has been demonstrated that explainable AI methods can enhance trust and transparency within automated diagnostic systems as they are combined with deep-learners. Indicatively, researchers have rendered neural network identified tumour regions with Grad-CAM to enable clinicians to confirm that the model concentrates on medically important aspects. The other research used a framework that was explainable and incorporated segmentation and classification models and used them to analyse MRI scans. Visual explanations and tumour predictions were produced by the system that could help clinicians gain a clearer understanding of how the model worked. Moreover, recent studies contrasting the variants of U-Net with explainability methods proved that Grad-cam
visualisation offers useful information about the parts of MRI images that the model utilises during the tumour segmentation.

**Multimodal MRI Analysis**

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

The diagnosis of brain tumour is often based on the application of two or more MRI modalities each with its specific image qualities of tissues. The T1 -weighted pictures can give you structural data about the parenchyma of the brain whereas the T2 -weighted pictures can emphasize fluid buildup. FLAIR sequences reduce cerebrospinal fluid signals, thereby enhancing tumour visibility. Multimodal models Deep-learning models using multimodal MRI inputs can extract complementary information in multiple imaging sequences. The method increases the accuracy of tumour segmentation and contributes to the differentiation of tumour and normal tissues of the surrounding brain. The empirical research that has been carried out has repeatedly shown that multimodal MRI analysis provides much higher brain tumour detection rates than single-modality analysis method.

**Brain Tumour Detection Research Trends in the Recent Times.**

It is noted that the main research trends in brain tumour detection are: - Transformer-based approaches to relationship capturing in global images. - Hybrid CNN-Transformer networks to enhance features in a model. - Unsupervised learning of reducing reliance on labelled data. - Interpretable AI models to enhance model transparency. - Minimal operating systems of real-time clinical implementation. These innovations are meant to enhance accuracy and reliability of the automated brain tumour detection systems.

Detection of brain tumor using medical imaging has emerged as one of the principal scientific investigations due to the growing rate of neurological conditions and the growing possibilities of the medical imaging technology. The Magnetic Resonance Imaging (MRI) technique, especially, provides an extremely high level of soft-tissue contrast, which allows clinicians to outline the pathological areas with an unprecedented level of granularity. the manual analysis of MRI images is tough and prone to get errors from observer ; therefore, this area is converted and shifted to fully automated, data-driven methods involving both machine learning and deep learning to improve accuracy in diagnostic tests and efficiency in the work process [1], [2].

Initial work in automated tumor detection consisted of traditional image-processing algorithms including thresholding and region growing and clustering and edge detection algorithms intended to single out tumor regions by means of intensity contrasts and geometric features in MRI images. These techniques produced early evidence of concept but were often weak in nature and could not be capable of accommodating intricate morphology that is typical of malignant lesions. Furthermore, the handcrafted dependence on the domain knowledge usually hampered generalizing to the various imaging cohorts that appeared in practice [3]. In order to overcome these drawbacks, researchers have turned to the use of supervised machine-learning classifiers, including Support Vector Machines (SVM), Decision Trees, Random Forests and k-nearest Neighbour (KNN) models, which are trained on manually-constructed descriptors, like texture, shape and intensity statistics obtained using MRI volumes. Experiments on comparative studies have shown that these algorithms were capable of outperforming straight image-processing baselines. However, they were limited in their ability to automatically acquire the complex, high-dimensional structures inherent to pathological imaging data because they relied on manual feature engineering [4], [5].

With the emergence of deep learning, medical analysis of images has been radically changed. Convolutional Neural Networks (CNNs) are recently at the state-of-the-art on a continuum of computer-vision problems: the key ones being object detection, image classification, and segmentation: through the automatic finding of hierarchical feature maps based on raw pixel values. This representational force has prompted the intensive use of CNN-based approaches to detecting and classifying brain tumors in publicly accessible MRI samples [6], [7]. The different CNN backbones such as VGGNet, AlexNet, ResNet, DenseNet, and EfficientNet have been used to distinguish between tumorous and non-tumorous voxel in MRI images. The use of residual networks (ResNet) in particular has attracted such interest because of their ability to run significantly deep architectures without the vanishing gradients. In practice an empirical evidence suggests that the classification accuracies of such models can reach up to over ninety percent of benchmark MRI corpora [8], [9].

Although CNNs are sufficiently useful in classification, clinical practice requires tumour boundaries to be localised accurately in three-dimensional imaging volumes. This has led to the creation of complex segmentation architecture that can mark the extent of tumour at the voxel level. The U-Net network which was presented as an innovative architecture of segmenting biomedical images is based on the
encoder-decoder architecture where the encoder progressively acquires feature representations at initially lower spatial resolutions (aggregation of features) and the decoder gradually learns to synthesize the finer spatial details to generate high-resolution segmentation maps [10]. One design capability that contributes to the success of U-Net is its skip-connection design, in which encoder layer maps bypass decoder layers with volumes of identical features. These shortcuts maintain fines grain spatial details that would otherwise be lost due to down-

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

sampling and thus the model is able to faithfully recreate fine anatomical structures. Therefore, U-Net has emerged as an omnipresent backbone of biomedical segmentation, such as the delineation of brain tumors, organ localization and lesion localization, thanks to its beautiful tradeoff between depth and spatial fidelity [11]. Although it can be successfully applied to the 2D image segmentation, the naturally three-dimensional character of MRI data restricts the use of planes-wise U-Net by ignoring the context in the following cuts. To address the same deficiency, scholars have suggested the 3D U-Net architecture that builds on the original design by substituting the two-dimensional convolutions with the volumetric convolutions. This volumetric extension allows the model to consume and run whole MRI volumes, providing spatial constraints between more than two slices and hence providing more coherent and anatomically consistent segmentation solutions [12].

The 3D U-Net models have repeatedly demonstrated a higher performance in segmenting brain tumours than 2D based, especially because they have the ability to question the entire volumetric structure of tumours. With these models offering the ability to bring the network into contact with spatially adjacent slices, this artificially provides greater capability to tell tumours apart of adjoining neural tissue, which has been demonstrated through models based on the Brain Tumor Segmentation (BraTS) dataset where 3D U-Nets achieve high levels of Dice similarity indices and relative segmentation fidelity compared to traditional 2-D frameworks [13]. In addition to the classic 3D U -Net architecture, a collection of advanced designs has appeared, designed to drive the performance of segmentation at its limits. Residual U-Nets add shortcut correlations that allow extended back-propagation paths, thus boosting the depth of representations.

U-Nets have gating schemes that select salient parts as the foreground background is suppressed unlike Dense U-Nets where features are propagated using a dense inter-layer connectivity topology which actively promotes feature reuse and gradient concentration throughout the network [14], [15]. Another confounding factor of detection efficacy is the quality of datasets available. This has been enabled by the BraTS challenge, which has made available a thoroughly curated multimodal MRI corpus of acquisitions which contains T1, T1 -contrast (T1c), T2, FLAIR, along with expert-labeled tumoral localization. These benchmarked resources provide the researchers with a reproducible platform to assess segmentation algorithms based on such metrics as Dice coefficient, sensitivity, specificity, and Intersection over Union (IoU) [16].

Although deep learning models have significant advances in the accuracy of prediction, the models remain opaque and therefore do not support clinical adoption. Deep neural networks are black-box and therefore it is hard to explain what is going on inside the computer as it is being predicted. Medical decision individuals on the other hand require systems that do not only make a prediction but also provide explanations, which has driven the development of Explainable Artificial Intelligence (XAI) systems that attempt to make deep learning processes transparent [17]. GradientWeighted Class Activation Mapping (Grad-CAM) has become popular as one of the XAI tools due to its capability of generating heatmaps onto the input images to highlight the parts that produce the most impact on a classification task. Grad-CAM may be used to detect tumor localisation in MRI images in the context of tumor detection, providing clinicians with the option of determining whether the model is directing attention where it should to clinical anatomy [18]. Another popular method is the Local Interpretable Model-Agnostic Explanations (LIME), an approximation method that calculates the output of a complex model locally by a model that is simple and interpretable. LIME can be used to examine the impact of discrete image patches on the final prediction by perturbing the input data and comparing the resulting change in predictions, which allows learning increasingly about the decision hierarchy of a model [19].

Based on the cooperative game theory, Shapley Additive Explanations (SHAP) provides a theoretically sound, coherent way of allocating the contribution to prediction based on the contribution of individual features. The SHAP scores can be applied to the medical imaging, where they associate the MRI voxels that contribute to the tumour detection the most, which supports the reliability and interpretability of the system [20]. Recent studies are moving towards the direction of combining these explainability methods with more advanced architectures like the 3D U-Net. Researchers can binarily identify tumour regions and understand the rationale behind each prediction with XAI modules directly integrated into segmentation pipelines, and this dual function is the most critical step towards inspiring clinical confidence in AI-enhanced diagnostics.

However, the discipline addresses long-standing predicaments. One of the bottlenecks is the lack of highly annotated medical datasets: the labeling of high quality requires the participation of skilled radiologists, which makes the process data collection resource-consuming and time-consuming. This results in most gain access to deep learning models having been trained on small sample sizes and are thus more susceptible to overfitting and a lack of generalisability. The second barrier is the huge computational costs of volumetric network training, which require large memory footprints, and computing capabilities, which could inhibit their use in clinics with time constraints. With light-weight architecture and model compression algorithms, the community is actively

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

developing and investigating alternative ways of reducing these limitations. In short the literature shows clearly that deep learning has significantly improved the speed of brain tumour detection and segmentation, acceleration in the accuracy of these processes. Headers to U-Net, especially when applied in three dimensions, have recorded impressive achievements in medical imaging.

Simultaneously, the interpretability gap has been mitigated by XAI tools, including Grad-CAM, LIME, and SHAP, which provide visual stories that can be used to explain algorithmic decision-making to clinicians. However, data (lack) and the computational overhead (time) and transparency considerations remain as high-priority research challenges. An effective remedy to the mentioned gaps by a conciliating approach of 3D U-Net segmentation and explainable AI is the proposed project, and the long-term objective is to present a suitable, precise, and clinically deployable brain tumour detection system.

A detailed attention was given to academic studies that utilized 3D U-Net models in delineation of brain tumours. The models were critically assessed with regard to their ability to consume volumetric MRI images, as well as define spatial dependencies among serial slices of images. The accompanying advantages and shortcomings of these architectures were analyzed at length.[39]. In addition to the segmentation paradigm, explainable artificial intelligence methodologies were also covered in the review.

Methods like Grad-CAM and SHAP were evaluated in terms of the effectiveness in improving the transparency of these models and providing readable details of AI-driven insights into the predictions generated by them [40]. The usefulness of the methods in the medical imaging field was proven by the exemplary case studies based on the available literature. After that, a theoretical framework is described that integrates the 3D U-Net segmentation stream with explainable AI practices. The

proposed model aims to address the failure flaw of existing methodologies through integrating accurate tumour localisation with clear model justifications.[42]. Such a combination promises to increase the reliability and the clinical usability of AI based brain tumour sensing systems.

U-Net with its variations is one of the best deep learning structures used in segmentation tasks with consistent good results. They have reported Dice scores exceeding 90 0 -percent with Tumour segmentation using U-Net-based models. Nonetheless, recent models like attention-based networks and transformer-based models have shown additional gains in the precision of segmentation. The attention mechanisms allow the model to focus on the important regions of the tumours and disregard the background noises. Explainable AI methods are important concerning the assessment of the credibility of these models. Visualization software like Grad-CAM has the capability of elucidating the predicates of the model by showing the areas of MRI images that the model relies on. According to the comparative analysis of the latest works, it seems that models that combine 3D segmentation architectures and explainability frameworks are the most appropriate ones in terms of accuracy and explainability. This has come with new challenges with the growing complexity of deep learning models.

Among the most serious challenges, there is the fact that deep learning predictions are not interpretable. Medical practitioners need clear definitions of diagnostic actions, especially when diagnostic actions are related to life-threatening situations like brain tumours. Clinicians do not trust black-box AI systems which generate predictions without explanations and restrict their use in healthcare settings. Explainable Artificial Intelligence (XAI) is a promising method of resolving such a problem. The XAI methods seek to offer an understanding of the internal decision-making mechanism of the machine learning models. Efforts like Grad-CAM produce visual heatmaps indicating areas within an image that contribute when making predictions with the model against the model. Likewise,SHAP values determine the importance of individual features to the output of the model. Such methods enable the clinicians to know the manner in which AI models perceive the medical images and can also test the accuracy of predictions. The high complexity of deep learning models is another limitation of brain tumour detection research because their computation cost is prohibitive.

The computationally resources needed to train large neural networks and data will be huge and are not necessarily accessible in most healthcare facilities. Scientists are consequently examining less

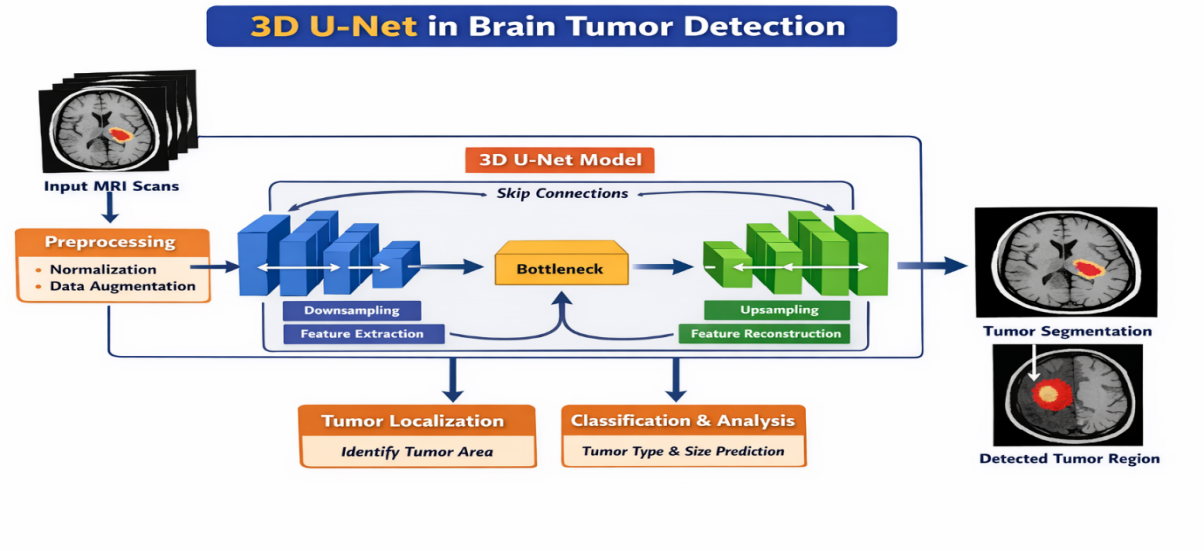
efficient structures and optimisation strategies to curtail the number of calculations. Moreover, the scarcity of various medical imaging data presents a major challenge towards coming up with more powerful AI models. The BraTS dataset is extremely important in numerous studies, yet it was found not to be as diverse a sample of clinical imaging data as it could be. Future studies should be dedicated to the development of bigger and more varied datasets to enhance generalisation of models. The aim of this review paper is to examine the latest advancements in detection of brain tumours with the help of 3 -D U-Net segmentation models and explainable AI methods. This paper will offer informative findings into future research pathways to create a more accurate, interpretable, and practical brain tumour detection device by analysing current literature and pinpointing major limitations. Once the first set of the research articles is collected, the first type of filtering was introduced to see

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

the most relevant studies.

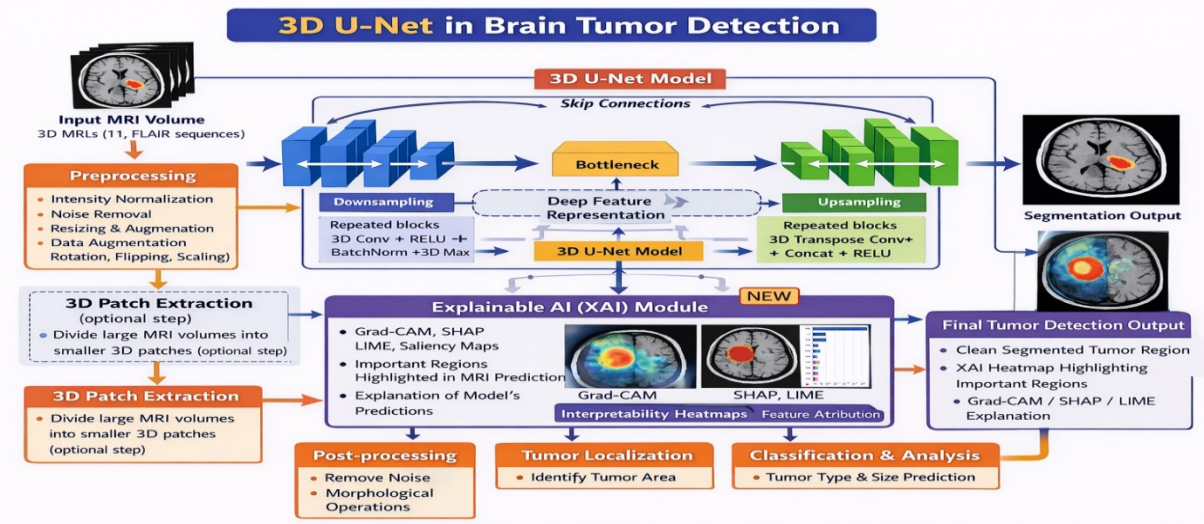
## METHODOLOGY

A detailed attention was given to academic studies that utilized 3D U-Net models in delineation of brain tumours. The models were critically assessed with regard to their ability to consume volumetric MRI images, as well as define spatial dependencies among serial slices of images. The accompanying advantages and shortcomings of these architectures were analyzed at length.[39].




In addition to the segmentation paradigm, explainable artificial intelligence methodologies were also covered in the review.

Methods like Grad-CAM and SHAP were evaluated in terms of the effectiveness in improving the transparency of these models and providing readable details of AI-driven insights into the predictions generated by them [40]. The usefulness of the methods in the medical imaging field was proven by the exemplary case studies based on the available literature. After that, a theoretical framework is described that integrates the 3D U-Net segmentation stream with explainable AI practices.




The proposed model aims to address the failure flaw of existing methodologies through integrating accurate tumour localisation with clear model justifications.[42]. Such a combination promises to increase the reliability and the clinical usability of AI based brain tumour sensing systems.

U-Net with its variations is one of the best deep learning structures used in segmentation tasks with consistent

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

good results. They have reported Dice scores exceeding 90 0 -percent with Tumour segmentation using U-Net-based models. Nonetheless, recent models like attention-based networks and transformer-based models have shown additional gains in the precision of segmentation. The attention mechanisms allow the model to focus on the important regions of the tumours and disregard the background noises. Explainable AI methods are important concerning the assessment of the credibility of these models. Visualization software like Grad-CAM has the capability of elucidating the predicates of the model by showing the areas of MRI images that the model relies on. According to the comparative analysis of the latest works, it seems that models that combine 3D segmentation architectures and explainability frameworks are the most appropriate ones in terms of accuracy and explainability. This has come with new challenges with the growing complexity of deep learning models.

Among the most serious challenges, there is the fact that deep learning predictions are not interpretable. Medical practitioners need clear definitions of diagnostic actions, especially when diagnostic actions are related to life-threatening situations like brain tumours. Clinicians do not trust black-box AI systems which generate predictions without explanations and restrict their use in healthcare settings. Explainable Artificial Intelligence (XAI) is a promising method of resolving such a problem. The XAI methods seek to offer an understanding of the internal decision-making mechanism of the machine learning models. Efforts like Grad-CAM produce visual heatmaps indicating areas within an image that contribute when making predictions with the model against the model. Likewise,SHAP values determine the importance of individual features to the output of the model. Such methods enable the clinicians to know the manner in which AI models perceive the medical images and can also test the accuracy of predictions. The high complexity of deep learning models is another limitation of brain tumour detection research because their computation cost is prohibitive.

The computationally resources needed to train large neural networks and data will be huge and are not necessarily accessible in most healthcare facilities. Scientists are consequently examining less

efficient structures and optimisation strategies to curtail the number of calculations. Moreover, the scarcity of various medical imaging data presents a major challenge towards coming up with more powerful AI models. The BraTS dataset is extremely important in numerous studies, yet it was found not to be as diverse a sample of clinical imaging data as it could be. Future studies should be dedicated to the development of bigger and more varied datasets to enhance generalisation of models. The aim of this review paper is to examine the latest advancements in detection of brain tumours with the help of 3 -D U-Net segmentation models and explainable AI methods. This paper will offer informative findings into future research pathways to create a more accurate, interpretable, and practical brain tumour detection device by analysing current literature and pinpointing major limitations. Once the first set of the research articles is collected, the first type of filtering was introduced to see the most relevant studies.

## RESULTS AND DISCUSSION

The next step in research on brain tumor detection should be creating more effective and decipherable deep-learning models. The major areas for contribution of research will be in the evolution of light deep-learning models that can be used in a clinical context to deploy real-time. The incorporation of multimodal medical data, which involves MRI, PET and genomic data, to enhance predictive attributes. The establishment of massive, heterogeneous medical-imaging data that capture heterogeneity of clinical populations.The development of explainable-AI methods that are specially designed to achieve medical image segmentation tasks.The application of AI systems to support real-time diagnostics being incorporated smoothly into hospital information infrastructures.

## ACKNOWLEDGEMENT

I would like to thank Prof. Y. R. Tayade who has guided, supported, and encouraged me during the work in great value. I am also grateful to my co-guide, Prof. N. M. Sapate who provided valuable suggestions and support. I am very greatful for getting the support from our Head of Department of Artificial Intelligence and Data Science Respected Dr. S. R. Zanwar Sir and their experience was very instrumental in the achievement of this project. I would also wish to thank all the faculty members and peers who helped me in this research.

## CONCLUSION

Detection of brain tumours with the help of deep learning methods has come to stand out as a highly developing field of research in the medical image analysis field. It has also been established that automated diagnostic systems have a potential capability of greatly improving the accuracy along with the efficiency of tumour detection as compared to the traditional manual procedures. In this review paper, recent contributions to the field of detecting brain tumours using the tools of the 3D U-Net segmentation and of explainable artificial

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

intelligence (XAI) are under scrutiny and clarified.
The research measures the advantages and limitations of extant deep-learning architectures and highlights the great importance of integrating XAI models to enhance model transparency. The 3D U-Net models have been shown to be at the forefront of segmentation of brain tumours through successful utilisation of volumetric information obtained off magnetic resonance images (MRI) scans. Such models are superior to the classic two dimensional segmentation methods because even the spatial dependencies between adjacent slices of the image are preserved.
Despite their achievements, deep-learning models are still faced with several problems to overcome such as severe computational complexity, limited interpretability, and lack of diversity in data sets. The explainable AI methods can also provide an interesting solution to the problem of model transparency, as they produce pictorial explanations of model predictions. These methods enable clinicians to conclude that AI models are focusing on the biologically relevant tumour locales. This intersection of the models of segmentation and explainable AI methods has the potential to significantly increase the accuracy and clinical usability of automated brain tumour visualisation technologies.

## REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” MICCAI, 2015.
[2] F. Milletari, N. Navab, and S. A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” 3DV, 2016.
[3] O. Çiçek et al., “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” MICCAI, 2016.
[4] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Communications of the ACM, 2017.
[5] K. He et al., “Deep Residual Learning for Image Recognition,” CVPR, 2016.
[6] O. Oktay et al., “Attention U-Net: Learning Where to Look for the Pancreas,” 2018.
[7] Z. Zhou et al., “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” IEEE TMI, 2020.
[8] J. Long et al., “Fully Convolutional Networks for Semantic Segmentation,” CVPR, 2015.
[9] B. H. Menze et al., “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” IEEE TMI, 2015.
[10] S. Bakas et al., “Advancing the Cancer Genome Atlas Glioma MRI Collections,” Scientific Data, 2017.
[11] S. Pereira et al., “Brain Tumor Segmentation Using CNNs in MRI Images,” IEEE TMI, 2016.
[12] K. Kamnitsas et al., “Efficient Multi Scale 3D CNN with CRF for Brain Lesion Segmentation,” Medical Image Analysis, 2017.
[13] H. Dong et al., “Automatic Brain Tumor Detection Using U Net Based FCN,” 2017.
[14] A. Casamitjana et al., “3D CNN Architectures for Brain Tumor Segmentation,” 2017.
[15] A. Beers et al., “Sequential 3D U Nets for Brain Tumor Segmentation,” 2017.
[16] X. Zhao et al., “FCNN and CRF Model for Brain Tumor Segmentation,” 2017.
[17] S. Montaha et al., “Brain Tumor Segmentation from 3D MRI Scans Using U Net,” SN Computer Science, 2023.
[18] P. Zheng et al., “Improved U Net for Brain Tumor Segmentation,” BMC Medical Imaging, 2022.
[19] A. Pourmahboubi et al., “Brain Tumor Segmentation Enhancement Using U Net,” BMC Medical Imaging, 2025.
[20] S. Swathi et al., “3D U Net for Brain Tumor Detection and Segmentation,” 2022.
[21] V. Rajinikanth et al., “3D MRI Segmentation Using U Net for Brain Tumor Detection,” Procedia Computer Science, 2023.
[22] Y. Zhang et al., “GenU Net++ for Brain Tumor Segmentation,” Symmetry, 2021.
[23] G. Litjens et al., “A Survey on Deep Learning in Medical Image Analysis,” Medical Image Analysis, 2017.
[24] A. Esteva et al., “A Guide to Deep Learning in Healthcare,” Nature Medicine, 2019.
[25] J. Ker et al., “Deep Learning Applications in Medical Image Analysis,” IEEE Access, 2018.
[26] H. Greenspan et al., “Deep Learning in Medical Imaging,” IEEE TMI, 2016.
[27] S. Lundervold and A. Lundervold, “An Overview of Deep Learning in Medical Imaging,” 2019.
[28] J. Shen et al., “Deep Learning in Medical Image Analysis,” Annual Review of Biomedical Engineering, 2017.
[29] R. Selvaraju et al., “Grad-CAM: Visual Explanations from Deep Networks,” ICCV, 2017.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

[30] S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," NeurIPS, 2017.
[31] M. Ribeiro et al., "Why Should I Trust You? Explaining the Predictions of Any Classifier," KDD, 2016.
[32] B. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence," IEEE TNNLS, 2021.
[33] W. Samek et al., "Explainable Artificial Intelligence: Interpreting Deep Learning Models," IEEE Signal Processing Magazine, 2017.
[34] A. Holzinger et al., "Explainable AI for the Medical Domain," 2017.
[35] A. Dosovitskiy et al., "Vision Transformer," ICLR, 2021.
[36] H. Cao et al., "Swin UNet: Transformer for Medical Image Segmentation," 2021.
[37] S. Hatamizadeh et al., "UNETR: Transformers for 3D Medical Image Segmentation," WACV, 2022.
[38] Z. Chen et al., "TransUNet: Transformers for Medical Image Segmentation," 2021.
[39] N. Do et al., "Brain Tumor Classification Using Deep Learning," IEEE Access, 2019.
[40] M. Havaei et al., "Brain Tumor Segmentation with Deep Neural Networks," Medical Image Analysis, 2017.
[41] A. Afshar et al., "Capsule Networks for Brain Tumor Classification," ICIP, 2019.
[42] S. Rehman et al., "Brain Tumor Detection Using CNN," IEEE Access, 2020.
[43] A. Amin et al., "Brain Tumor Detection Using Deep Learning," IEEE Access, 2020.
[44] A. Khan et al., "Survey of Deep Learning for Brain Tumor Detection," IEEE Access, 2021.
[45] S. Albahli et al., "Explainable AI Framework for Brain Tumor Detection," IEEE Access, 2022.
[46] Y. LeCun et al., "Deep Learning," Nature, 2015.
[47] T. Litjens et al., "Computer-Aided Detection in Medical Imaging," IEEE TMI, 2017.
[48] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," Neural Networks, 2015.
[49] I. Goodfellow et al., "Deep Learning," MIT Press, 2016.
[50] D. Shen et al., "Deep Learning in Medical Image Analysis," Annual Review of Biomedical Engineering, 2017.