2024 International Conference on Recent Innovation in Smart and Sustainable Technology (ICRISST) | 979-8-3503-0914-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICRISST59181.2024.10921899

3D SEGMENTATION OF MRI USING MULTIMODAL FOR DETECTION OF BRAIN TUMOUR

Sanjay D
Department of Computer Science andEngineering
Sathyabama Institute of Science andTechnology
Chennai, Tamilnadu, India.
sanjaydilli9@gmail.com

V Devarsh
Department of Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, Tamilnadu, India.
vdevarsh2016@gmail.com

E Murali*
Department of Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, Tamilnadu, India.
emurali88@gmail.com

J Cruz Antony
Department of Computer Science andEngineering
Sathyabama Institute of Science andTechnology
Chennai, Tamilnadu, India.
jcruzantony@gmail.com

R.Vignesh
Department of Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, Tamilnadu, India.
vignesh.cse@sathyabama.ac.in

Abstract— **Determining brain tumors from MRI can be challenging due to several factors, and it typically requires expertise in neuroimaging and collaboration between radiologists, neurologists, and other specialists. Some tumors may have a subtle or a typical appearance on MRI, making them difficult to distinguish from normal brain tissue. Brain tumors can be highly heterogeneous in terms of silhouette, scope, andinternal structure. Tumors can have different types of tissues, such as necrotic and enhancing regions, making it challenging to accurately delineate their boundaries. Brain tumors can be highly heterogeneous in terms of silhouette, scope, andinternal structure. Tumors can have different types of tissues, such as necrotic, cystic, or enhancing regions, making itchallenging to accurately delineate their boundaries. Deep learning has shown promising results in various medical imaging tasks, including the identification and classification of brain tumors, such as gliomas, from MRI images. In this paper, an attempt to design an ensemble model for the prediction of3D brain tumor segmentation and performance is going to be evaluated using deep learning models such as UNET, FPN, and UNET+FPN. The above models are evaluated on the BraTS 2021 (Brain Tumour Segmentation) dataset, and comparing themetrics of each model reveals that UNET and, an enhanced model that combines UNET and FPN, UNET+FPN has the highest accuracy of 0.9959 and 0.9954 respectively.**

***Keywords*— UNET, FPN, Ensemble Model**

I. INTRODUCTION

Brain tumor segmentation is a critical task in medical image analysis that plays a pivotal role in the diagnosis, treatment planning, and monitoring of patients with brain tumors. Accurate delineation of tumor boundaries from magnetic resonance imaging (MRI) scans is essential for assessing tumor size, location, and progression. Magnetic resonance imaging (MRI) features are as regarded one of the preferable approaches to identify [1] a brain tumor because it produces a more detailed and clearer picture than CT scans. Different contrast images generated by multimodal MRI procedures give complementary information that aids in segmenting the brain tumor and its surrounding tissues, which is crucial in the diagnosis and treatment of brain tumors. Deep learning methods, particularly Convolutional Neural Networks (CNNs), have shown impressive achievements across a range of image analysis assignments, including the segmentation of medical images. CNNs are adept at acquiring hierarchical features from data, allowing them to grasp complex patterns and spatial correlations within images. In the context of brain tumor segmentation, deep learning models can learn to distinguish between healthy and pathological tissue, as well as different tumor subtypes. The goal of segmentation is to modify the way different parts of a picture are characterized, making it simpler to comprehend sections of the image with different attributes. After dividing the brain image into thus many different pieces, each region becomes spatially continuous. Traditionally, this segmentation was done manually by trained professionals, a time-consuming and subjective process prone to errors. Brain tumors exhibit diverse morphological characteristics, and their precise segmentation from medical images is challenging due to variations in shape, size, and intensity. Consequently, the automated segmentation of brain MRI images can significantly enhance diagnostic and treatment procedures, particularly in situations where there is a shortage of radiologists and skilled professionals. However, precise tumor segmentation remains difficult due to glioma variability in terms of shape, size, and appearance, as well as the unclear and fuzzy boundary that exists between cancer and brain tissue [2]. The MRI data's intensity fluctuation adds to the difficulties. As a result, it is still subject to development, allowing for more research into improved segmentation approaches and accuracy. In this study, image segmentation models such as UNET, FPN, and UNET+FPN for confining and detecting the brain tumor from the 3D MRI images. The main objectives of this study are:

- The automatic segmentation method for brain tumors utilizes informative image slices extracted from a 3D multimodal MRI volume, aiming to decrease computational time while enhancing segmentation accuracy.
- Comparative analysis of image segmentation models based on the evaluation metrics for brain segmentation
- A hybrid model combined above mentioned models using an average ensemble technique and study of its performance on test data

II. RELATED WORKS

This paper attempts to design an automated AI-based brain tumor from an MRI scan. The study aims to develop a deep-learning model specifically for brain tumor segmentation. The authors review existing segmentation models and highlight the popularity of the 3D U-Net model. The proposed model is similar to the 3D U-Net architecture and consists of 6 encoding and decoding layers. It is trained on the BRATS dataset using dice loss and focal loss. The model achieves high accuracy, loss, and IOU scores. Comparative experiments indicate that the proposed model surpasses alternative models. Ablation studies are carried out to assess the influence of various layers. In summary, the proposed model attains state-of-the-art performance in brain tumor segmentation [5].

The paper focuses on semantic image segmentation, assigning class labels to pixels in an image for scene understanding. Existing methods are accurate but too slow for real-time applications like self-driving cars. The authors propose LinkNet, a CNN architecture that achieves real-time performance while maintaining accuracy. LinkNet uses an encoder-decoder structure, with the encoder based on pre-trained ResNet-18 and the decoder upsampling features to full resolution. Shortcut connections preserve spatial information. The authors trained LinkNet on CamVid and Cityscapes datasets, using cross-entropy loss, data augmentation, and class weighting [4].

"BMC Medical Imaging" introduces an algorithm proposal aimed at enhancing the U-Net network and refining the FPN strategy, along with data augmentation techniques, to accurately segment the brain tumor region [7]. Although more experiments may be needed to validate the precise effectiveness of this approach, the paper showcases an improvement in the segmented brain tumor region utilizing the proposed U-Net network with FPN and data enhancement techniques [8]. To offer valid MRI scan results, deep learning approaches to merge deep neural networks with a CNN model. The CNN layout that was demonstrated was based on a three-layer fully connected neural network [12] [13]. The F-score achieved 97.33%, with an accuracy of 96.05% [14]. A 3D CNN architecture was devised for tumor extraction, incorporating transfer learning for classification, yielding an accuracy of 98.32% on the BRATS 2015 dataset [15]. Based on pyramid scene parsing (PSP) for semantic segmentation. The proposed model uses convolutional neural networks (CNN) and pyramid pooling modules to extract spatial information from images. Overall, the paper mainly focuses on the development, evaluation, and visualization of the proposed PSPNet architecture for semantic segmentation [6]. The authors of reference [11] have solved a binary classification difficulty with concerning MRI imaging of brain malignancies. They used recurrent feature elimination (RFE) after extracting features using ALEXnet and VGG16. To fulfill the categorization problem, they eventually deployed a Support Vector Machine (SVM), which achieved an overall accuracy of 96%.[9]. The author employed superpixel algorithms for tumor segmentation and identification subsequent to transfer learning. The tumor was divided into two categories using the superpixel approach. The average dice index that resulted from this was 0.93 when compared to the ground truth data.

III. METHODOLOGY

The main objective of this study is to develop a brain tumor segmentation algorithm capable of segmenting the four sub-tumor sections across all modalities. This is achieved by comparing existing deep learning models and employing an ensemble approach, wherein predictions from each model are combined using the average ensemble technique.The dataset used for this study is BraTS 2021, which consists of 1250 image files each image file comprises multimodal MRI scans, including T1- weighted, T2-weighted with contrast enhancement, T2- weighted, and fluid-attenuated inversion recovery (FLAIR) image, each image as the shape of (128,128,155) These different modalities provide complementary information about the brain and tumor tissue. The overall workflow of training models is mentioned in Fig.1.

A. Data Preprocessing

The dataset is divided into training, validation, and testing sets. The MRI images FLAIR, T1-weighted (T1ce) are used as input data(X) and segmented used for ground truth image (y). Each image is reshaped and sliced with an image size of

128. One-hot encoding on the segmentation labels. It converts the segmentation data into a one-hot encoded format with four classes.

B. Training

Each model was trained on 30 epochs using 300 training images and 250 validation images per epoch. During model training, the callbacks function is utilized; the early callbacks function is used to verify validation accuracy with patiences of 5 in order to stop training if there is no progress in validation accuracy after 5 epochs. The categorical cross entropy loss is used as a loss function. Adam optimizer implemented, with a learning rate of 0.001.

C. Evaluation Metrics

Accuracy, Mean Intersection Over Union, Sensitivity, Dice Coefficient, and Specificity were all used to assess the suggested model's performance [10]. These metrics can be expressed as TP, FP, TN, and FN, representing the counts of true positives, false positives, true negatives, and false negatives, respectively, using their respective acronyms. These metrics are used for evaluating image segmentation whereas other metrics

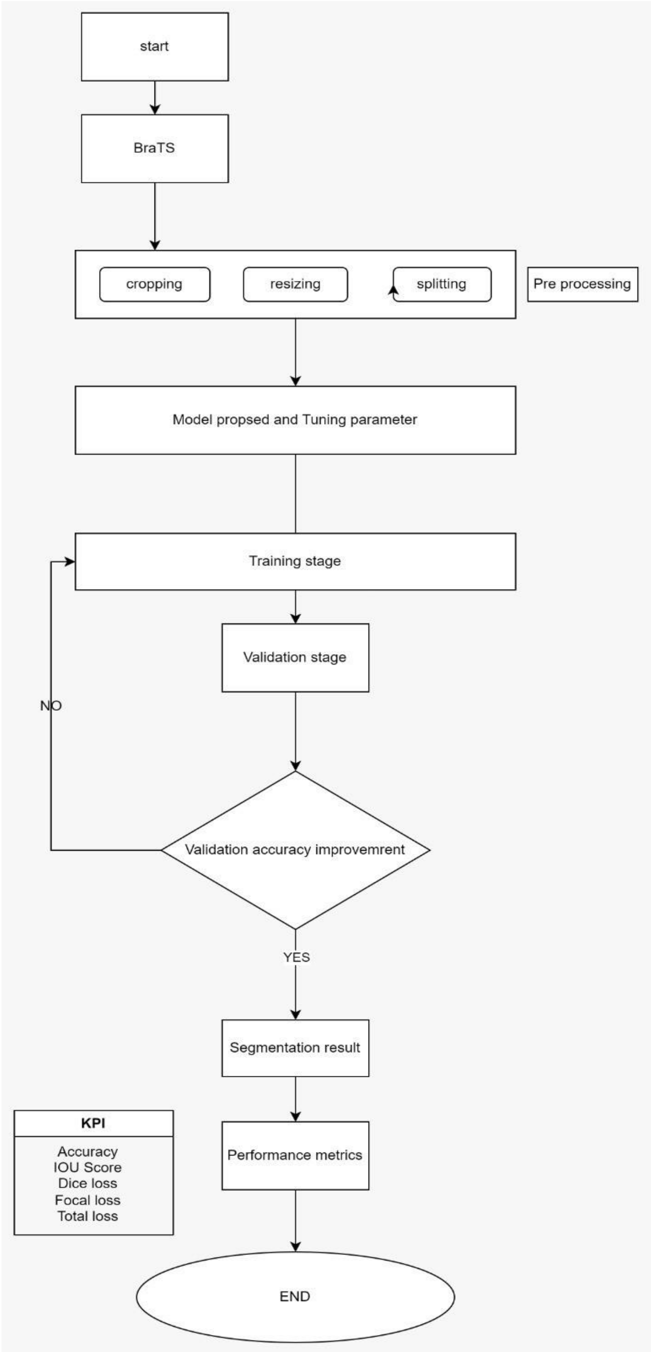

Fig.1. The overall workflow of training the models

are not compatible and are only used for binary classification. The metrics are defined as follows

a) Intersection Over Union is measured as the ratio of the overlap area to the entire area covered by both the predicted and ground truth areas. It quantifies how closely the anticipated region matches the ground truth region.

$$IOU = \frac{TP}{TP+FP+FN} \qquad (1)$$

b) The dice coefficient is a computation of an overlap-based metric that measures the spatial overlap between the ground truth mask and forecast mask.

$$DC = \frac{2 * target \cap prediction}{target + prediction} \qquad (2)$$

c) Specificity is the proportion of real negative values that were accurately detected.

$$Specivicity = \frac{TN}{TN+FP} \qquad (3)$$

d) Sensitivity is the proportion of real positive values that were accurately detected.

$$Sensitivity = \frac{TP}{TP+FN} \qquad (4)$$

e) Accuracy is a probability statistic that evaluates the rate to which segmentation results suit the base true segmentation mask.

$$Accuracy = \frac{TN+TP}{TN+TP+FP+FN} \qquad (5)$$

IV. RESULT

In this part, we will go over the experimental findings, network architecture for all of the models, model parameters, training details, and segmentation outcomes. In last, a comprehensive comparison of projected deep learning models is done using several assessment measures.

A. UNET

Ronneberger et al. introduced UNET, an architecture for semantic segmentation that employs a Fully Convolution Network Model [3]. The contrading path and expanding path together make up the model contrading path has 5 encoders. The first encoder has two convolution layers witha kernel size of 3x3. The remaining encoders have max pooling with a kernel size of 2x2 and two convolution layers with a kernel size of 3x3. The expanding path has four encoders each encoder has Three convolution layers where one layer has a kernel sign of 2 and the remaining layer has a kernel size of 3. Each convolution has a Relu function and the kernel initializer. The middle layer of the Model has a dropout of 0.5. The output layer of the model has a convolution of 1x1 to split each component feature vector into a relevant number of classes

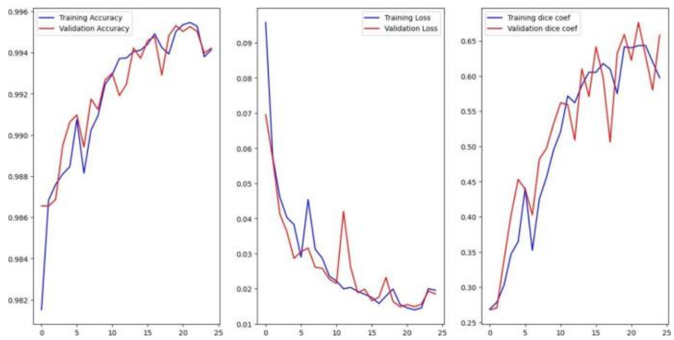

Fig.2. Graph for training log of UNET

On seeing the graph Fig. 2 It is evident that the model's accuracy improves with each epoch of training. Additionally, the validation accuracy consistently surpasses the training accuracy. This suggests that the model is effectively generalizing to new data and is not overfitting to the training data. The loss graph certainly demonstrates that the model's loss decreases with each epoch of training. Furthermore, the validation loss consistently remains lower than the training loss. This indicates that the model is successfully minimizing the error between predicted and actual values and is not overfitting to the training data.

B. FPN

FPN addresses the need for precise localization and strong semantic features by constructing a pyramid of feature maps. To achieve this, FPN utilizes two main techniques were used Firstly, the top-down pathway involves upsampling high-resolution but semantically weak features from shallower layers and combining them with lower-resolution but semantically strong features from deeper layers. FPN employs lateral connections, which directly add high-resolution features from shallower layers to their corresponding counterparts in the pyramid. To construct the pyramid, FPN adds 1x1 convolutional layers on top of each high-level element map from the CNN. These layers bring the features to the same resolution as the high-resolution feature maps by upsampling. The upsampled features are then directly added to their corresponding counterparts at each level of the pyramid through lateral connections.

In the top-down pathway, the upsampled features are element-wise summed with the corresponding lower-level features in the pyramid. This integration of information from different scales and resolutions enhances the ability of FPN to detect objects. The FPN architecture in the bottom-to-up path is composed of four convolutions. These convolutions have a respective number of filters [256, 512, 1024, 2056] and strides of [4, 8, 16, 32], with a kernel size of 3x3.On the other hand, the top-to-down path has four conditions, each with 256 filters and a kernel size of 1x1. Additionally, there are three samplings: up1, up2, and up3. Each sampling has

256 filters and kernel sizes of 2x2, 2x2, and 8x8 respectively. The up samplings up1, up2, and up3 are combined with their nearest neighbors in the bottom-to-up path using the Add() skip connection.

Furthermore, the model includes a decoder path consisting of four decoders. Each decoder has two convolutions and samplings. The convolutions have 128 filters and a kernel size of [128, 64, 32] to recover the spatial information from each Add() layer.

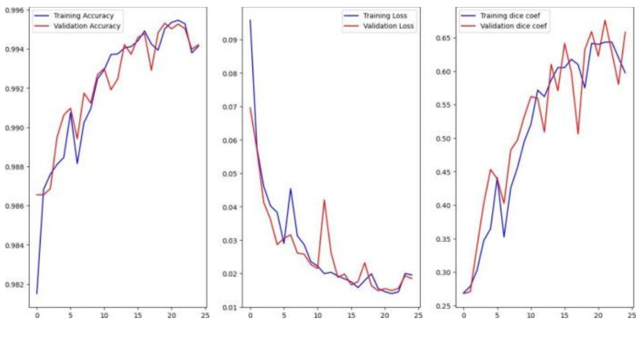

Fig.3. Graph for training log of FPN

The output of the up-sampling is concatenated into a single layer. Then, a convolution with a softmax activation function is applied to generate the final output.

The graph depicted in Fig. 3. illustrates a progressive increase in both training and validation accuracies over time, suggesting effective learning and enhancement in the model's predictive abilities. Both training Loss and validation Loss decreaseover time, indicating an improvement in model training. training dice coefficient and validation dice coefficient increase over time, indicating an improvement in the model's segmentation capability.

C. UNET+FPN

Upon comparing the architecture of FPN and U-Net, it becomes evident that they share similarities. The horizontal connection present in FPN can be achieved through the horizontal connection in U-Net. This feature enables the expansion of the U-Net model using the FPN structure. By leveraging the U-Net structure to its fullest potential, the U-Net model effectively utilizes information from various scales. In the middle section of the encoder and decoder components of the UNET, we implemented a bottom-to-top pathway of the FPN. This pathway is responsible for restoring information from lower resolution to higher resolution. The FPN pathway consists of four upsamplings with a kernel size of (2,2) and four 1x1 convolutions with filter sizes of [32,64,128,256]. Each upsampling operation is accompanied by skip connections that connect the encoder and decoder of the UNET.

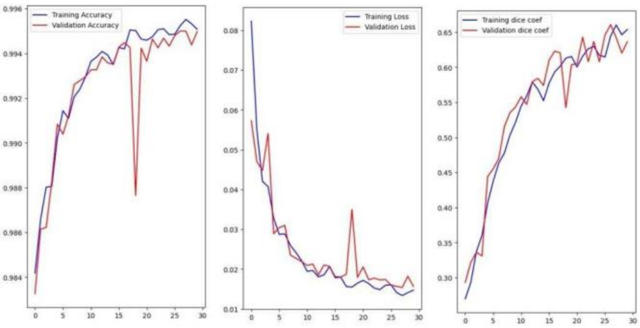

Fig.4. Graph for training log of UNET + FPN

The graph Fig. 4 illustrates that, In the beginning, both accuracies rise rapidly, but after around epoch 10, they reach a plateau, suggesting that the model is effectively learning but starting to stabilize its performance.

Initially, the training loss decreases and then stabilizes. On the other hand, the validation loss experiences a spike before decreasing and stabilizing as well. This pattern indicates a potential case of overfitting at the beginning, where the model is overly fitted to the training data. On the other hand, the graph on the right side presents the dice coefficients for the training and validation datasets throughout the epochs. Higher values in the graph indicate improved performance in terms of similarity between the predicted and actual results in tasks like image segmentation. As time progresses, both coefficients gradually increase but with some fluctuations, suggesting that the model is continuously learning and making adjustments.

D. ENSEMBLE MODEL

In the end, developed an ensemble model consisting of five sub-models: UNET, FPN, and UNET+FPN. This ensemble model combines the outputs of these sub-models to generate a single output. The input layer of the ensemble model is set to (128,128,2). The average layer calculates the average of the outputs from the sub-models.

E. Comparison of Evaluation Metrics of Models

Fig. 5 shows that models UNET+FPN, and Ensemble model have the lowest loss, suggesting superior model performance. Every model has similar accuracy with slight differences. In comparison the models, UNET, UNET+FPN have the greatest Mean IoU score. Except for FPN with slight differences, all models in the research had the same accuracy and sensitivity score. Evaluation Metrics of modelsare mentioned in Table 1.

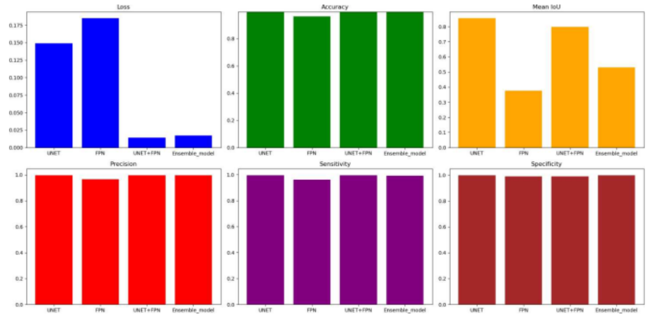

Fig.5. Comparison of Evaluation Metrics of Models

TABLE I. Evaluation Metrics of Models

Metrics	UNET	FPN	UNET+FPN	ENSEMBLE MODEL
LOSS	0.149	0.0622	0.0144	0.0171
ACCURACY	0.9959	0.9929	0.9953	0.9945
MEAN IOU	0.8538	0.3750	0.7981	0.5308
PRECISION	0.9963	0.9948	0.9967	0.9961
SENSITIVITY	0.9955	0.9913	0.9942	0.9933
SPECIFICITY	0.9988	0.9983	0.9889	0.9987

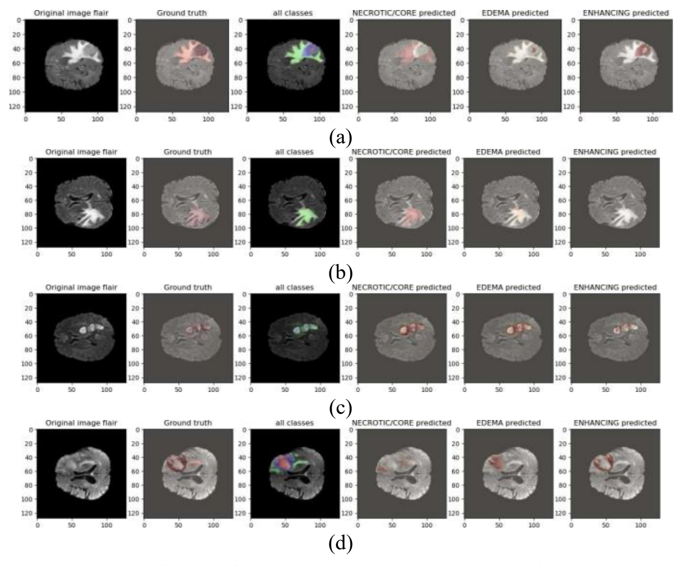

Fig.6. Prediction of models on test image (a)UNET, (b) FPN, (c) UNET+FPN, (d) Ensemble model

V. CONCLUSION

The main objective of this study is to evaluate the effectiveness of various deep learning models for image segmentation, specifically for the rapid identification of brain tumors. The process of segmenting medical images with the assistance of radiologists can be extremely time-consuming and resource-intensive, making it impractical in remote areas. Therefore, the implementation of advanced automation techniques for brain tumor detection can greatly benefit a significant number of cases.

In order to address this issue, various models were utilized in this article, namely UNET, FPN, and UNET+FPN. Each model demonstrated different levels of accuracy during the data testing phase which is shown in Fig. 6. Additionally, an ensemble model was created by averaging the predictions from each individual model[16]. The model UNET, UNET+FPN exhibited the highest accuracy among all models, while the UNET+FPN greater accuracy model followed closely behind.FPN and Ensemble model achieved stable performance

As a result, physicians are able to strategize treatment and monitor the progression of cancers during the diagnostic stage. The particular region of interest has been split into segments, and the existence of the tumor can be identified through an extremely accurate model. This method offers the benefit of enhancing image segmentation and spatial localization to a greater extent, resulting in superior performance compared to previous systems. Furthermore, it requires less computational time and can be trained at a faster pace compared to networks with fewer parameters.

In the forthcoming studies, the emphasis will be on employing diverse classifier techniques to enhance precision while reducing error rates. Additionally, this approach could be tailored to predict the prognosis of individuals afflicted with brain tumors. In future research, the utilization of larger and more varied datasets may prove beneficial in conductingtests under real-world circumstances and clinical trials.

REFERENCES

[1] S. Das, S. Bose, G. K. Nayak, S. C. Satapathy, and S. Saxena, “Brain tumor segmentation and overall survival period prediction in glioblastoma multiforme using radiomic features,” Concurrency Comput.: Pract. Experience, p. e6501.

[2] M. B. Naceur, R. Saouli, M. Akil and R. Kachouri, "Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in MRI images", Comput. Methods Programs Biomed., vol. 166, pp. 39-49, Nov. 2018.

[3] Ronneberger, O., Fischer, P., Brox, T (2015). “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: Medical Image Computing and Computer-Assisted Intervention. pp. 234–241 (2015).

[4] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 2017 , pp. 1 - 4 , doi: 10 . 1109 / VCIP.2017.8305148

//www.sciencedirect.com/science/article/pii/S095741741930586X

[5] Kajal M., & Mittal, A. (2022). A modified U-Net based architecture for brain tumour segmentation on BRATS 2020

[6] Zhao, M., Xin, J., Wang, Z., Wang, X., & Wang, Z. (2022). Interpretable Model Based on Pyramid Scene Parsing Features for Brain Tumor MRI Image Segmentation. Computational and Mathematical Methods in Medicine, *2022*.

[7] Smarta Sangui, Tamim Iqbal, Piyush Chandra Chandra, Swarup Kr Ghosh, Anupam Ghosh,3D MRI Segmentation using U-Net Architecture for the detection of Brain Tumor, Procedia Computer Science, Volume 218, 2023, Pages 542-553, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2023.01.036.

[8] Sun, H., Yang, S., Chen, L., Liao, P., Liu, X., Liu, Y., & Wang, N. (2023). Brain tumor image segmentation based on improved FPN. *BMC Medical Imaging*, *23*(1), 172

[9] Mesut T, Zafer C, Burhan E, (2020). "Classification of Brain MRI Using Hyper Column Technique with Convolutional Neural Network and Feature Selection Method". Expert Systems with Applications. 149

[10] S. Roy, R. Saha, S. Sarkar, R. Mehera, R. K. Pal and S. K. Bandyopadhyay, "Brain Tumour Segmentation Using S-Net and SA- Net," in *IEEE Access*, vol. 11, pp. 28658-28679, 2023, doi: 10.1109/ ACCESS.2023.3257722

[11] Ahuja S, Panigrahi BK, Gandhi T (2020). "Transfer Learning Based Brain Tumour Detection and Segmentation using Superpixel Technique," 2020 International Conference on Contemporary Computing and Applications (IC3A), 2020, pp. 244 – 249

[12] Choudhury CL, Mahanty C, Kumar R, Mishra BK, (2020). "Brain Tumour Detection and Classification Using Convolutional NeuralNetwork and Deep Neural Network," 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), pp. 1-4.

[13] Ghosh SK, Biswas B, Ghosh A (2020). "Restoration of Mammograms by Using Deep Convolutional Denoising Auto-Encoders". In:Behera, H., Nayak, J., Naik, B., Pelusi, D. (eds) Computational Intelligence in Data Mining. Advances in Intelligent Systems and Computing, vol 990 . Springer ,Singapore.

[14] Rehman A, Khan MA, Saba T, Mehmood Z, Tariq U, Ayesha N (2021). "Microscopic brain tumor detection and classification using 3D CNN and feature selection architecture". Microsc Res Tech. 84: 133– 149.

[15] Tripathi P, Singh VK, Trivedi MC (2021). "Brain tumor segmentation in magnetic resonance imaging using OKM approach", Materials Today: Proceedings, Vol. 37, pp 1334-1340.

[16] Abdullah, A. ., Murali, E., S, S. ., Balusamy, B. ., & Rajashree, S. (2023). Design of Multiple Ontology Based Agro Knowledge Mining Model. International Journal on Recent and Innovation Trends in Computing and Communication, 11(7), 47–56. https://doi.org/10.17762/ijritcc.v11i7.7829