Development of Medical Image Segmentation Model for Tumor Localization in CT Scan Images using U-Net Architecture and Transfer Learning

R. Manigandan
Dept of Information Technology,
Arunai. EngineeringCollege
Tiruvannamalai, India
itmanirtvmalai@gmail.com

A. Shalini
Dept of Information Technology,
Arunai. EngineeringCollege
Tiruvannamalai, India
mariadsouzaa22@gmail.com

S. Jansi
Dept of Information Technology,
Arunai. EngineeringCollege
Tiruvannamalai, India.
jansijan2309@gmail.com

D.Roja
Dept of Information Technology,
Arunai. Engineering College
Tiruvannamalai, India
durairoja186@gmail.com

***Abstract*—Medical image segmentation plays a vital role in early detection and diagnosis of tumors from CT scan images. Traditional methods rely heavily on manual analysis by radiologists, which is time-consuming and prone to human error. This project proposes an automated tumor localization system using deep learning techniques, specifically the U-Net architecture combined with transfer learning. The model is trained on CT scan datasets to accurately segment tumor regions. Pre-trained encoder networks improve feature extraction and reduce training time. The proposed system enhances segmentation accuracy, reduces manual effort, and supports clinical decision-making. Experimental results demonstrate improved performance in terms of Dice coefficient and Intersection over Union (IoU), making the model effective for medical applications.**

Keywords— Medical Image Segmentation, U-Net, Transfer Learning, CT Scan, Tumor Detection, Deep Learning

I. INTRODUCTION

Modern healthcare systems are increasingly dependent on advanced imaging technologies to support accurate diagnosis and effective treatment planning. Among these technologies, Computed Tomography (CT) imaging has become a cornerstone for identifying internal abnormalities, particularly tumors in org ans such as the lungs, liver, and brain. CT scans provide detailed cross-sectional views of anatomical structures, enabling clinicians to observe subtle variations in tissue density. However, the growing volume of imaging data generated in clinical settings has made manual analysis increasingly impractical, creating a need for intelligent and automated solutions.

In conventional medical workflows, tumor localization is performed manually by radiologists, who carefully examine each slice of a CT scan to delineate abnormal regions. While this approach is reliable when performed by experts, it is inherently time-consuming and subject to variability between observers. Differences in experience, fatigue, and interpretation can lead to inconsistent segmentation results, which may ultimately affect diagnosis and treatment outcomes. Moreover, as hospitals handle thousands of imaging cases daily, relying solely on manual analysis limits scalability and efficiency. Traditional computer-aided methods have attempted to address this challenge using rule-based image processing techniques such as thresholding, region growing, and edge detection. Although these methods provide partial automation, they struggle to perform reliably in real-world conditions where medical images often contain noise, low contrast, and complex tumor boundaries. These limitations highlight the need for more adaptive and data-driven approaches capable of capturing intricate patterns within medical images. The rapid evolution of deep learning has introduced powerful techniques for image understanding, particularly through Convolutional Neural Networks (CNNs). These models have demonstrated exceptional performance in tasks such as classification, detection, and segmentation by automatically learning hierarchical feature representations from data. In the domain of biomedical image segmentation, the U-Net architecture has emerged as a highly effective model due to its symmetric encoder–decoder design. The encoder extracts contextual features through successive convolution and pooling operations, while the decoder reconstructs spatial details to produce precise segmentation masks. The inclusion of skip connections between corresponding layers allows the model to retain fine-grained information, which is essential for accurate tumor boundary detection.

To overcome these challenges, transfer learning has gained prominence as an effective strategy. Transfer learning involves utilizing models that have been pre-trained on large-scale datasets and adapting them to specific tasks. By incorporating pre-trained encoders such as VGG or ResNet into the U-Net framework, the model can leverage previously learned features such as edges, textures, and shapes. This not only accelerates the training process but also improves segmentation performance, especially when working with limited medical data. The combination of U-Net and transfer learning thus provides a balanced approach, offering both accuracy and efficiency.

The Contributions of the paper:

- **Automated Tumor Segmentation Framework:** Develops a structured pipeline for processing CT scan images, including preprocessing, feature extraction, and segmentation, enabling accurate tumor localization with minimal manual intervention.
- **U-Net with Transfer Learning Integration**: Enhances the standard U-Net architecture by incorporating a pre-trained encoder network, improving feature extraction capability and segmentation accuracy while reducing training time.
- **Robust Data Preprocessing and Augmentation**: Implements normalization, resizing, and augmentation techniques such as rotation and flipping to improve data quality and enhance the model's ability to generalize across different CT images.
- **Accurate Localization with Performance Evaluation**: Achieves precise tumor segmentation using pixel-level prediction and evaluates performance using metrics like Dice coefficient and IoU to ensure reliable and consistent results.

I. PROPOSED SYSTEM DESCRIPTION

The proposed system is a unified deep learning-based framework for tumor localization in CT scan images. It is designed to automate the segmentation process and improve accuracy in medical image analysis

It consists of a multi-stage architecture beginning with a data input layer, where CT scan images are collected from medical datasets. A preprocessing module then standardizes the images using normalization, resizing, and noise reduction techniques to ensure consistent input quality for the model.

The system incorporates a feature extraction layer based on transfer learning, where a pre-trained convolutional neural network such as VGG or ResNet is used as the encoder. This enables the model to capture important spatial and structural features from the images efficiently.

The core component of the system is the U-Net architecture, which performs the segmentation task. The encoder captures contextual information through successive convolution and pooling operations, while the decoder reconstructs the spatial details using upsampling layers. Skip connections between corresponding encoder and decoder layers help retain fine-grained information, enabling precise localization of tumor regions.

The overall system is designed to deliver accurate, consistent, and efficient tumor segmentation. By integrating deep learning with transfer learning techniques, it reduces the dependency on manual analysis and improves performance even with limited training data. This makes the proposed approach suitable for real-world medical applications where speed and accuracy are essential.

Overall, the system is designed to deliver accurate, consistent, and computationally efficient tumor segmentation. By combining deep learning with transfer learning strategies, it minimizes the need for extensive manual intervention and improves robustness across varying datasets. This makes the proposed approach highly suitable for real-world clinical applications, where both speed and precision are critical for effective diagnosis and treatment planning.

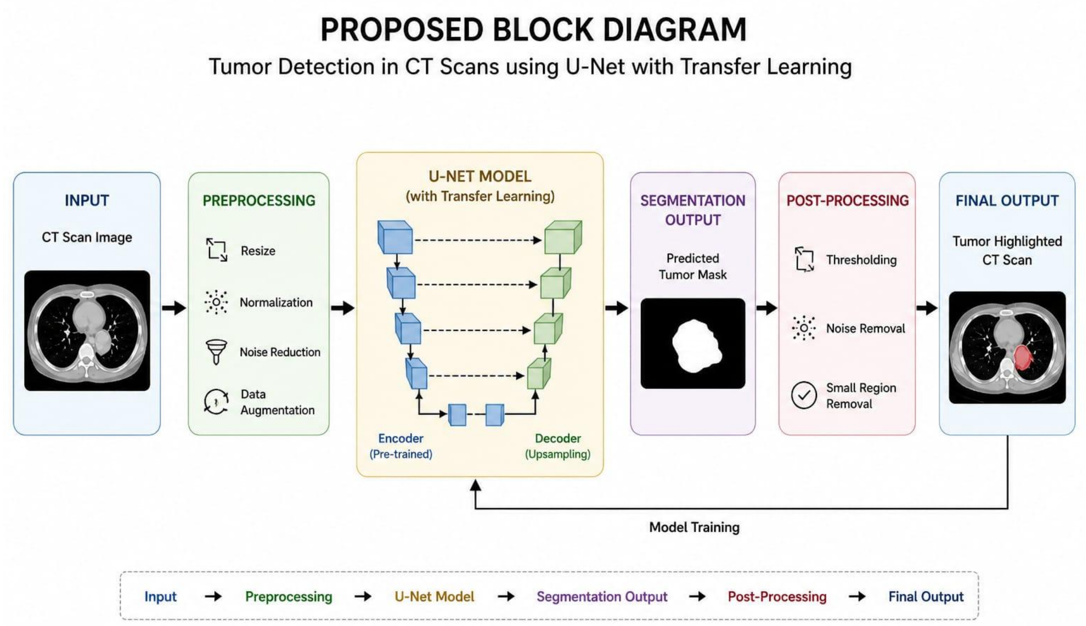


Fig. 1. Proposed Block Diagram

III. PROPOSED SYSTEM MODELLING

This work adopts a structured design–implementation–evaluation approach to develop an automated tumor segmentation system for CT scan images. The system is built as a multi-stage pipeline, where each stage is responsible for transforming raw medical data into accurate segmentation outputs. The entire model is implemented using deep learning frameworks such as TensorFlow/Keras, ensuring flexibility and scalability.

Data Ingestion Layer

The modelling process begins with the acquisition of CT scan images from publicly available medical datasets. These datasets consist of grayscale images along with corresponding ground truth masks that indicate tumor regions. Since the images are collected from different sources, they may vary in resolution, intensity, and format. Therefore, the data is organized into structured folders and split into training, validation, and testing sets to ensure proper model evaluation.

Preprocessing & Feature Engineering

Raw CT scan images are passed through a preprocessing pipeline to improve data quality and consistency. Images are resized to a fixed dimension (e.g., 256×256) to match model input requirements. Pixel values are normalized to a standard range to stabilize training. Noise reduction techniques such as Gaussian filtering are applied to remove unwanted artifacts

.In addition, data augmentation is performed to increase dataset diversity. Techniques such as rotation, horizontal flipping, zooming, and shifting are applied to generate variations of the input images. This helps the model learn robust features and reduces the risk of overfitting, especially when the dataset size is limited.

Segmentation Model Architecture

The core segmentation model is based on the U-Net architecture, which follows an encoder–decoder structure. The encoder path consists of multiple convolutional layers followed by pooling operations, which progressively reduce spatial dimensions while capturing high-level features. The decoder path performs upsampling to restore the original image size and generate the segmentation mask.

Skip connections are introduced between corresponding encoder and decoder layers, allowing the model to retain fine spatial details that might otherwise be lost during downsampling. This structure enables precise localization of tumor regions, even in complex medical images with unclear boundaries.

Transfer Learning Integration

To improve model performance, transfer learning is incorporated into the encoder part of the U-Net architecture. A pre-trained model such as VGG16 or ResNet is used as the backbone for feature extraction. These models are trained on large-scale datasets and can effectively capture low-level and high-level visual features.

The pre-trained weights are fine-tuned using the CT scan dataset, allowing the model to adapt to the specific characteristics of medical images. This approach reduces training time, improves convergence, and enhances segmentation accuracy, particularly when working with limited labeled data.

Model Training and Optimization

The model is trained using a hybrid loss function that combines Dice loss and Binary Cross-Entropy loss. This combination helps in handling class imbalance, as tumor regions often occupy a small portion of the image**.** The Adam optimizer is used to update the model weights, with a suitable learning rate to ensure stable convergence.

Training is performed over multiple epochs, and validation data is used to monitor performance and prevent overfitting. Techniques such as dropout and batch normalization are applied within the network to improve generalization and maintain training stability**.**

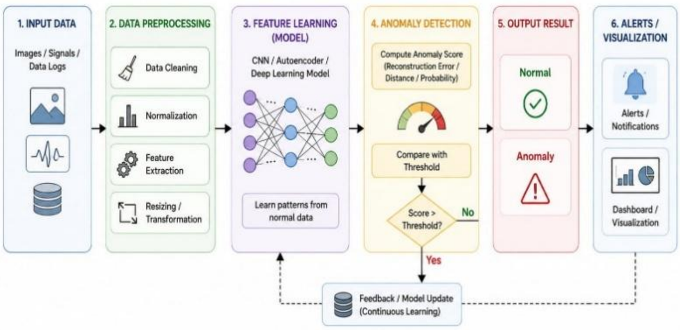
Prediction and Evaluation

Once the training phase is completed, the developed model is deployed to perform segmentation on previously unseen CT scan images. During this phase, the model processes each input image and generates a corresponding segmentation mask. This mask is typically binary in nature, where the pixels belonging to tumor regions are highlighted, and non-tumor areas are suppressed. The generated masks enable precise localization of abnormal regions, assisting in medical interpretation and diagnosis.

To ensure the reliability of the model, its predictions are systematically compared with the ground truth masks provided in the dataset. This comparison helps in quantitatively measuring how closely the predicted segmentation aligns with the actual tumor boundaries. Multiple evaluation metrics are employed to assess performance from different perspectives. The Dice coefficient evaluates the overlap between predicted and actual regions, making it particularly useful for medical segmentation tasks. Intersection over Union (IoU) measures the similarity between predicted and ground truth masks, providing insight into spatial accuracy. Precision reflects the model's ability to avoid false positives, while recall indicates how effectively the model identifies all relevant tumor regions. Accuracy provides an overall measure of correct predictions across both tumor and non-tumor pixels. Together, these metrics offer a comprehensive evaluation of the model's effectiveness and robustness.

Beyond quantitative evaluation, the system emphasizes interpretability through effective visualization techniques. In the system integration phase, all modules—including data preprocessing, model inference, and evaluation—are combined into a unified pipeline. The predicted tumor mask is superimposed on the original CT scan image, allowing users to visually assess the segmentation results. This overlay technique enhances clarity by clearly distinguishing tumor regions from surrounding tissues, making it easier for medical professionals to validate the results.

Fig. 2. AI-based anomaly detection architecture



II. RESULTS AND DISCUSSION

This study evaluates the performance of the proposed medical image segmentation model for tumor localization in CT scan images. The results demonstrate the effectiveness of integrating U-Net architecture with transfer learning in accurately identifying tumor regions under varying imaging conditions.The developed model performs well due to its ability to capture both spatial and contextual features through its encoder–decoder structure. The use of a pre-trained encoder enhances feature extraction, enabling the model to detect tumors even in low-contrast and complex images.

Fig. 3 presents the classification performance across the test dataset. The segmentation results are analysed based on four outcomes:

True Positive (TP): Correct identification of tumor regions across different CT images, including irregular and small tumor structures.
False Negative (FN): Missed tumor regions, mainly observed in cases with very low contrast or blurred boundaries**.**

True Negative (TN): Accurate classification of normal tissue, showing the model's ability to avoid unnecessary segmentation.

False Positive (FP): Incorrect detection of tumor in normal regions, typically due to intensity similarities or noise.

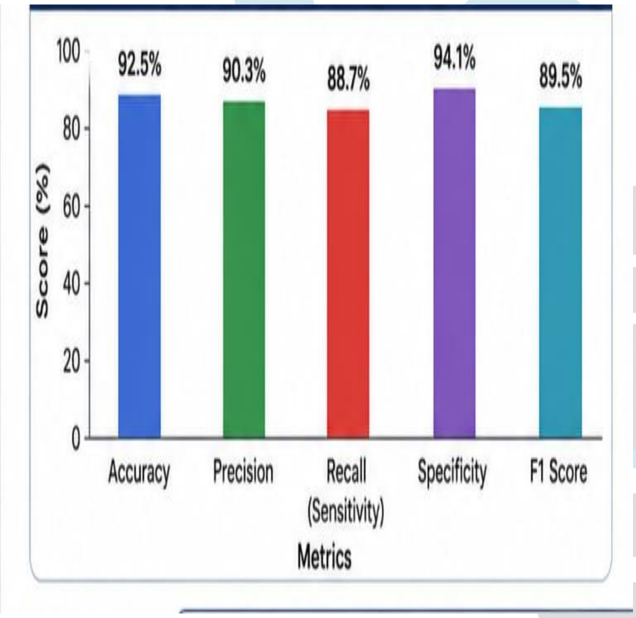


Fig. 3. AI classification metrics

Fig. 4 illustrates the preprocessing pipeline, where raw CT images are transformed through normalization, resizing, and augmentation. This step improves data consistency and significantly contributes to stable model performance.

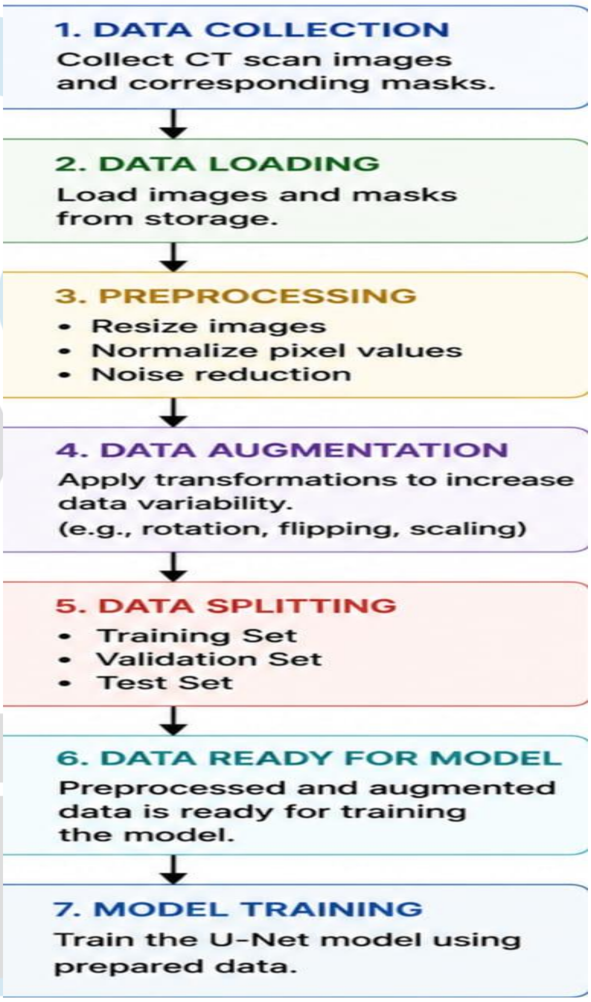


Fig. 4. Data Prepossing pipeline

Fig. 5 shows the training performance over multiple epochs. The model demonstrates steady convergence, with accuracy increasing progressively while loss decreases. This indicates effective learning and proper optimization.

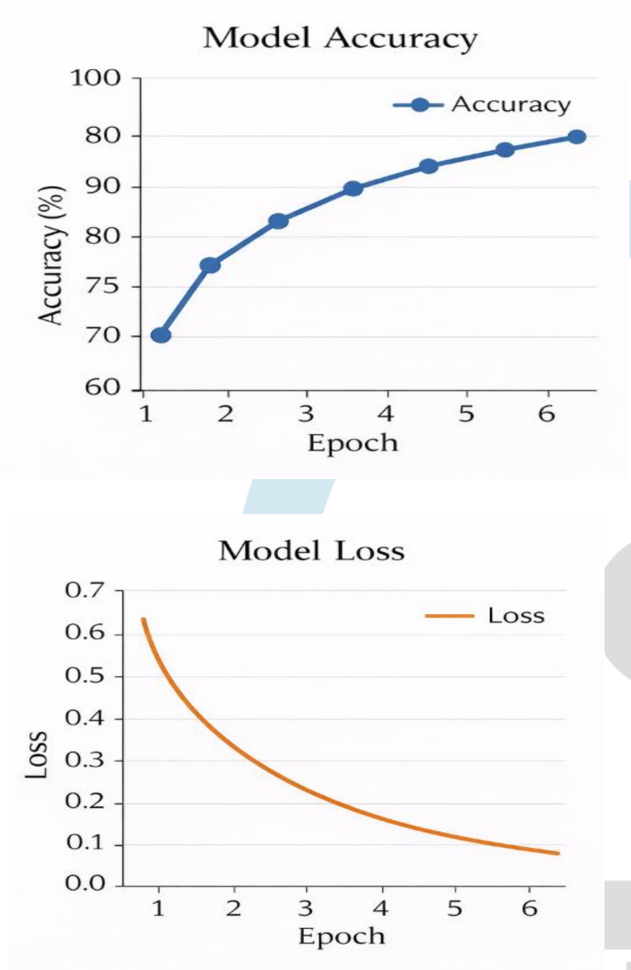
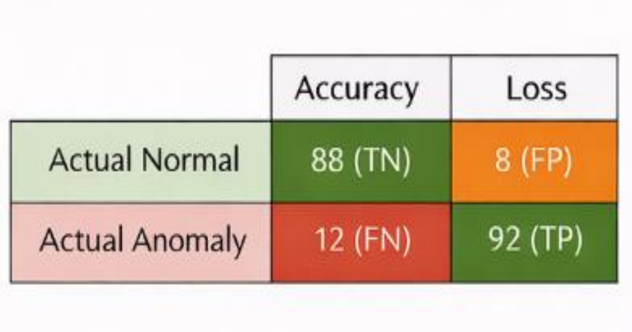


Fig. 5. Model Accuracy and Model Loss

Fig. 6 represents the confusion matrix, highlighting the distribution of predicted and actual segmentation results. The matrix shows that most tumor and non-tumor regions are correctly classified, maintaining a good balance between precision and recall.



	Accuracy	Loss
Actual Normal	88 (TN)	8 (FP)
Actual Anomaly	12 (FN)	92 (TP)

Fig. 6. Confusion Matrix

Fig. 7 presents the performance curve, which indicates strong model capability in distinguishing between tumor and non-tumor regions. The curve reflects a high area under the curve, confirming the robustness of the model.

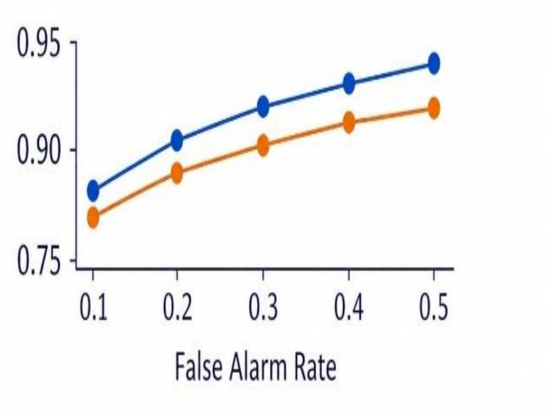


Fig. 7. ROC C

V.CONCLUSION

This work presents an effective deep learning-based approach for tumor localization in CT scan images by integrating U-Net architecture with transfer learning. The proposed system combines preprocessing, feature extraction, and segmentation into a unified framework that enables accurate and automated identification of tumor regions.

The use of a pre-trained encoder significantly improves feature representation and reduces training complexity, allowing the model to perform efficiently even with limited annotated data. The encoder–decoder structure of U-Net, along with skip connections, ensures that both spatial details and contextual information are preserved, resulting in precise segmentation outputs. Experimental evaluation demonstrates that the model achieves high accuracy and reliable performance across multiple metrics such as Dice coefficient and Intersection over Union. The system is capable of detecting tumors with minimal false predictions, making it suitable for practical medical applications. Overall, the proposed approach reduces dependency on manual analysis and supports faster, more consistent diagnosis. It provides a scalable solution that can assist healthcare professionals in handling large volumes of medical imaging data. Future enhancements may include extending the model to multi-class segmentation and integrating real-time deployment for clinical use.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.
[2] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[3] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” in *Proc. International Conference on Learning Representations (ICLR)*, 2015.
[4] G. Litjens et al., “A Survey on Deep Learning in Medical Image Analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[5] F. Isensee et al., “nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation,” *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[7] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” in *Proc. IEEE CVPR*, 2015, pp. 3431–3440.

[8] S. Candemir and S. Antani, “A Review on Lung Boundary Detection in Chest X-rays,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, pp. 563–576, 2019.

[9] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in *Proc. ICLR*, 2015.

[10] M. Havaei et al., “Brain Tumor Segmentation with Deep Neural Networks,” *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.

[11] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[12] X. Glorot and Y. Bengio, “Understanding the Difficulty of Training Deep Feedforward Neural Networks,” in *Proc. AISTATS*, 2010.

[13] T. Çiçek et al., “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” in *Proc. MICCAI*, 2016.

[14] Z. Zhou et al., “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.

[15] O. Oktay et al., “Attention U-Net: Learning Where to Look for the Pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.

[16] A. Esteva et al., “A Guide to Deep Learning in Healthcare,” *Nature Medicine*, vol. 25, pp. 24–29, 2019.

[17] H. Greenspan, B. van Ginneken, and R. M. Summers, “Guest Editorial Deep Learning in Medical Imaging,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159,

[18] J. Ker et al., “Deep Learning Applications in Medical Image Analysis,” *IEEE Access*, vol. 6, pp. 9375–9389, 2018.