

Mandibular Image Segmentation and 3D Reconstruction Using U-Net Model

Mambaul Izzi¹⁾, Chastine Fatichah^{2,*}, and Hadziq Fabroyir³⁾

^{1, 2, 3)} Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

E-mail: 6025231074@student.its.ac.id¹⁾, chastine@its.ac.id²⁾, and hadziq@its.ac.id³⁾

ABSTRACT

Segmentation of 3D models from Digital Imaging and Communications in Medicine (DICOM) images plays a crucial role in mandibular reconstructive surgery planning. This study segments DICOM images by modifying the slice images, preventing resampling. The main challenge lies in the low quality of DICOM images or noise, which affects segmentation and reconstruction accuracy. This study integrates a deep learning-based U-Net model for automatic segmentation to address this issue. Evaluation results show that U-Net achieves the highest segmentation accuracy with a Dice Similarity Coefficient (DSC) of 92.96%, precision of 95.90%, sensitivity of 93.22%, and specificity of 99.74%. Regarding computational efficiency, U-Net demonstrates the fastest training time at 755 minutes, compared to Res-U-Net at 785 minutes and Dense-U-Net at 767 minutes. These findings demonstrate that integrating segmentation with 3D reconstruction enhances mandibular anatomical visualization, improves surgical planning efficiency, and provides an interactive simulation tool for personalized treatment and professional training. Adopting DICOM standards ensures accessibility across medical devices and supports interoperability within healthcare systems.

Keywords: U-Net, automatic segmentation, jawbone, 3D reconstruction, DICOM, medical AI

1. Introduction

Mandibular reconstruction is a complex surgical procedure crucial in restoring function and facial aesthetics to patients with mandibular defects. Mandibular defects can result from various causes, such as trauma, tumor, infection, or congenital abnormalities. Restoration of the patient's quality of life and psychosocial aspects. In modern clinical practice, successful mandibular reconstruction relies heavily on precise pre-operative planning and an in-depth understanding of the patient's anatomy. 3D reconstruction improves the quality of training for medical personnel through interactive simulation, which facilitates more personalized care and ultimately results in more optimal outcomes [1][2]. While there have been advances in automated segmentation technologies, they often have difficulty accurately recognizing anatomical structures without manual intervention. Inaccurate segmentation can result in 3D models lacking precision, negatively affecting surgical outcomes. In addition, 3D reconstruction techniques have limitations that are influenced by the resolution of images generated by medical imaging. Low resolution can obscure important anatomical details indispensable for successful surgical planning. The quality of 3D reconstruction is highly dependent on the capabilities of the software and hardware used. Less sophisticated software or inadequate hardware may limit the quality of 3D visualization [3]. The quality of 3D reconstruction is highly dependent on the capabilities of the software and hardware used. Less sophisticated software or inadequate hardware can limit the quality of 3D visualization [4].

To overcome these limitations, one solution that is increasingly being applied is the U-Net segmentation technique [5]. The U-Net architecture adopts an Encoder-Decoder structure, which allows the network to learn complex features from medical images. The encoder part serves to capture the overall context, while the decoder incrementally increases the resolution so that the segmentation remains accurate even at the smallest details [6].

* Corresponding author.

Received: February 14th, 2024. Revised: January 11th, 2025. Accepted: January 31st, 2025.

Available online: February 25th, 2025.

© 2025 The Authors. This is an open access article under the CC BY-SA license (<https://creativecommons.org/licenses/by-sa/4.0/>)

DOI: 10.12962/j24068535.v23i1.a1245

U-Net is designed to produce better segmentation by reducing variations caused by human subjectivity. This architecture improves segmentation quality by utilizing full connectivity in the image, which reduces the possibility of segmentation errors that usually occur in traditional methods. After successful segmentation, the results can be used for more accurate 3D reconstruction of the tumor, which is crucial in surgical planning. U-Net can support 3D visualization of tumors with higher accuracy, which is important for therapeutic planning and surgical intervention [7]. In this study, segmentation is also performed using other models, such as Res U-Net and Dense U-Net, to evaluate the performance of various architectures in improving segmentation accuracy.

For the results of 3D segmentation and reconstruction to be optimally accessed and used across various medical platforms and devices, there is a need for consistent medical image data management and exchange standards. One of the most widely used standards in this field is Digital Imaging and Communications in Medicine (DICOM). DICOM is an international standard for storing, retrieving, and transmitting medical image information. DICOM enables the integration of varied medical equipment and health information systems, providing a consistent framework for various medical imaging applications [8]. Each DICOM file stores not only image data but also important metadata information such as slice thickness, distance between slices, image resolution, and patient clinical data, which supports more accurate handling and processing of image data. DICOM is designed to support various imaging devices, such as MRI, CT scan, ultrasound, and radiography, enabling data exchange and analysis across platforms and equipment. Surgeons can use 3D imaging built from DICOM slices plan procedures more accurately [9].

This research contributes to a 3D segmentation and reconstruction method using U-Net models to improve the accuracy and efficiency of the medical analysis process on the mandibular bone. With a workflow that includes dataset preparation, model building and training, and evaluation and refinement, the proposed method enables automatic and more precise segmentation than manual techniques. Integrating the resulting model with 3D reconstruction supports more effective visualization and surgical planning. The model evaluation and refinement process ensure that this solution can be implemented practically and according to clinical needs, providing real benefits in medical treatments and interventions.

2. Related Research

In the paper written by Sha et al [10] The focus is improving segmentation efficiency in medical image analysis using a three-dimensional (3D) to two-dimensional (2D) conversion approach through a 2D U-Net architecture. This method involves cutting a 3D image into 2D slices at specific intervals and applying maximum intensity projection to generate a 2D image. This approach reduces computational cost and utilizes a proven effective 2D segmentation technique. The advantage of this method is accelerating the image analysis process, which is essential for rapid diagnosis, especially in emergencies such as cerebrovascular incidents. However, a major drawback is the potential loss of volumetric information during conversion, which sometimes affects segmentation accuracy. Nevertheless, experiments show that this conversion approach is efficient without sacrificing quality.

Harnessing the power of U-Net architecture in medical image analysis is not limited to 2D approaches. The use of U-Net in brain tumor segmentation, as discussed next, demonstrates the potential of 3D methods to address more complex medical image analysis challenges. Similar research on brain tumor segmentation using 3D U-Net architecture focuses on improving segmentation accuracy through hyperparameter optimization. This segmentation is crucial for medical diagnosis and treatment, as manual segmentation is time-consuming and subjective. The 3D U-Net model was trained with brain MRI data from the 2018 BraTS challenge using three loss functions (Dice loss, Log-Cosh loss, and Focal Tversky loss) and three optimizers (RMSProp, Adam, and Nadam). The results showed that combining the Log-Cosh loss function with the RMSProp optimizer achieved the highest Dice coefficient of 0.75. Although this automated method reduces segmentation time and reliance on experts, hyperparameter adjustment is still required for optimal results [11]. A similar application of 3D U-Net [12] was explored for the automatic segmentation of mandibular and maxillary substructures in CT scans of head and neck cancer patients, achieving high accuracy while drastically reducing processing time. This work demonstrates how 3D U-Net can

support detailed anatomical analysis, which is essential for tasks such as radiation treatment planning and implant-based rehabilitation.

In addition to brain tumors, the U-Net architecture is also applied in segmenting other organs, such as the spleen, with adjustments tailored for different data characteristics. Research conducted by Fatma et al. [13] evaluated the performance of U-Net in 3D image segmentation, with a particular focus on spleen segmentation using the MSD dataset. The method involves splitting the 3D image into 2D chunks, training the chunks with U-Net, and recombining the results into a volumetric image. The best result of F1-Score 0.840 and 89% accuracy shows the effectiveness of this method in reducing computational burden without compromising segmentation accuracy. However, more complex 3D models require more resources and face challenges in handling variations in object size [13].

In addition to organs such as the spleen, the application of U-Net also shows superior performance in brain tumor segmentation, especially with an optimized architecture to handle complex morphological challenges. Other research [14] focuses on glioblastoma tumor segmentation using U-Net optimized with Deep Fully Convolutional Networks (D-FCNs). This approach is important as manual segmentation is time-consuming and requires high expertise. The method uses MRI multimodal fusion to combine information from different modalities, while intensity normalization and data augmentation are used to address data heterogeneity and prevent overfitting. With the modification of skip connections, the model enables better high-resolution information flow between the encoding and decoding layers, achieving a Dice Score of 0.88 for the entire tumor. It is shown that the U-Net DFCN architecture can provide accurate and efficient segmentation for different grades of glioma. Not only focusing on brain tumors, but deep learning approaches such as U-Net have also demonstrated broader capabilities in medical image processing and surpassed traditional methods with more efficient automation. The development of deep learning has revolutionized brain tumor segmentation by reducing the reliance on traditional methods that rely on man-made features. The CNN architecture and its variants, such as U-Net and FCN, enable the automatic detection of complex features. The BRATS dataset is used as an evaluation standard to measure the segmentation performance of various gliomas. However, although 3D CNN and ensemble techniques capture spatial information effectively, these methods require high computation. Preprocessing, such as intensity normalization and data augmentation, is key in overcoming limitations, including class imbalance and lack of annotation data [15].

In addition to its application in tumor segmentation, U-Net was also successfully used in other medical applications, such as lung segmentation, which highlights the flexibility of this architecture in various clinical contexts. Research on semantic lung segmentation with U-Net shows that this architecture can adapt to the diagnostic needs of lung diseases such as lung cancer and pneumonia. A modified model with batch normalization and dropout overcomes the challenge of overfitting while adding additional filters speeds up the training process. Tests on 240 images yielded an accuracy of 98.3% and a Dice Coefficient of 96.29%, showing superior performance compared to previous methods. However, the limitations of small datasets necessitate further research with larger datasets and transfer learning techniques for more optimized results [16].

Overall, these studies show that U-Net architecture, with its various modifications, has become a very effective tool in medical image segmentation. The U-Net segmentation technique, with its deep encoder-decoder network structure, has improved segmentation accuracy over traditional methods. This architecture reduces the time and variability of manual segmentation, enabling faster and more accurate surgical planning and diagnosis. The integration of DICOM standards supports better 3D visualization of slice data, and the application of U-Net in various medical applications in both 2D and 3D has strengthened computational efficiency. Going forward, further U-Net architecture and algorithm development are needed to make this technology more reliable and effective in clinical practice.

3. Research Methodology

The experimental design flow illustrated in Fig. 1 depicts the stages in the research applying deep learning for medical image segmentation. This workflow starts with a data preparation stage where the data to be processed

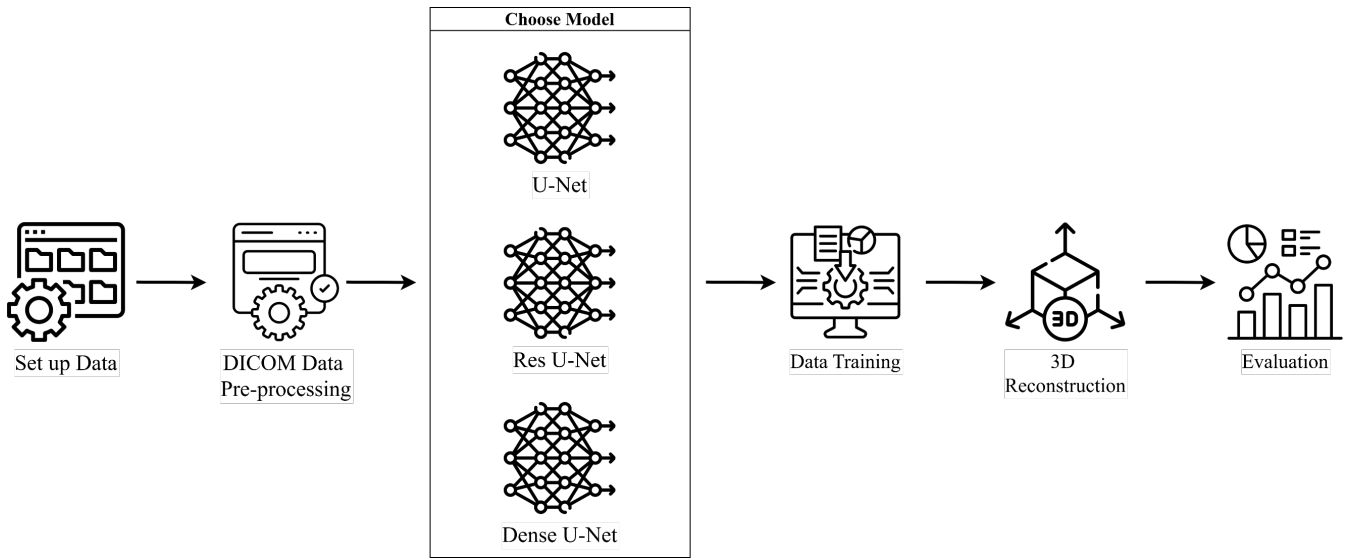


Fig. 1: Flowchart of experimental design.

is prepared in advance. The second stage, DICOM data pre-processing, involves processing Digital Imaging and Communications in Medicine (DICOM) data to prepare it as a model. In the model selection stage, three deep learning model architectures are compared: U-Net, Res U-Net, and Dense U-Net. After selecting the model, the process proceeds to the data training stage, where the selected model is trained using the pre-processed data. The next stage is “3D Reconstruction,” where the model output creates a 3D reconstruction of the segmented area. The last stage, “Evaluation,” assesses the model and segmentation results to measure effectiveness and accuracy in clinical applications. This workflow illustrates the integration between initial data processing, selection of an appropriate machine learning model, and model evaluation, which are important aspects of diagnostic applications in the medical world.

3.1. Model Segmentation

This study uses three models to segment DICOM data: U-Net, Res U-Net, and Dense U-Net models. These models were chosen to evaluate the effectiveness and accuracy of various deep-learning architectures in medical image segmentation, especially in the mandibular area.

A. U-Net

The U-Net architecture shown in Table 1 starts with an input layer that receives an image of size (512,512,1). The first layer performs an initial convolution using Conv2D with batch normalization and ReLU, resulting in an output of (512,512,32) with 320 parameters. Next, is a series of residual blocks starting with residual block 1, which processes the input into (512,512,32) using 9,248 parameters, followed by MaxPool2D for downsampling. The process continues through residual blocks 2 to 4, where each block increases the number of channels while decreasing the spatial dimension through MaxPool2D. The bridge/bottleneck layer processes data of size (32,32,256) to (32,32,512) using 1,180,160 parameters, being the deepest point of the U architecture. The decoder starts with upsampling block 1, which changes the dimension to (64,64,256) with 524,544 parameters. The decoder process proceeds through four stages, each consisting of upsampling followed by Conv2D with batch normalization and ReLU. Each stage gradually restores the spatial dimension while reducing the number of channels. Finally, the architecture ends with decoder block 4, which processes the input (512,512,64) into (512,512,32) using 18,464 parameters, and the last output layer uses Conv2D to produce a final output of size (512,512,1) with 33 parameters. The overall architecture exhibits a symmetrical structure between the encoder and decoder, which is the hallmark of the U-Net architecture. Each step in the model is carefully linked to ensure that local and contextual information is captured, thus making U-Net highly effective for medical image segmentation tasks [17].

Table 1: U-Net network architecture.

Layer	Output Shape	Output Shape	Parameter	Functions
Input	-	(512,512,1)	0	Input layer
Conv2D + BN + ReLU	(512,512,1)	(512,512,32)	320	First convolution layer
Conv2D + BN	(512,512,32)	(512,512,32)	9,248	Residual block 1
MaxPool2D	(512,512,32)	(256,256,32)	0	Downsampling
Conv2D + BN + ReLU	(256,256,32)	(256,256,64)	18,496	Residual block 2
MaxPool2D	(256,256,64)	(128,128,64)	0	Downsampling
Conv2D + BN + ReLU	(128,128,64)	(128,128,128)	73,856	Residual block 3
MaxPool2D	(128,128,128)	(64,64,128)	0	Downsampling
Conv2D + BN + ReLU	(64,64,128)	(64,64,256)	295,168	Residual block 4
MaxPool2D	(64,64,256)	(32,32,256)	0	Downsampling
Conv2D + BN + ReLU	(32,32,256)	(32,32,512)	1,180,160	Bridge/bottleneck
Upsampling	(32,32,512)	(64,64,256)	524,544	Upsampling block 1
Conv2D + BN + ReLU	(64,64,512)	(64,64,256)	1,179,904	Decoder block 1
Upsampling	(64,64,256)	(128,128,128)	131,200	Upsampling block 2
Conv2D + BN + ReLU	(128,128,256)	(128,128,128)	295,040	Decoder block 2
Upsampling	(128,128,128)	(256,256,64)	32,832	Upsampling block 3
Conv2D + BN + ReLU	(256,256,128)	(256,256,64)	73,792	Decoder block 3
Upsampling	(256,256,64)	(512,512,32)	8,224	Upsampling block 4
Conv2D + BN + ReLU	(512,512,64)	(512,512,32)	18,464	Decoder block 4
Conv2D	(512,512,32)	(512,512,1)	33	Output layer

B. Res U-Net

The ResUNet model shown is a complex convolutional neural network architecture for image segmentation tasks. The following is a detailed explanation in paragraph form: This architecture can be seen in Table 2, starting with an input layer that receives a 512x512 pixel image with one channel. Next, the network is divided into encoder and decoder paths connected through skip connections. In the encoder path, the input image goes through a series of convolution blocks consisting of Conv2D, BatchNormalization, and ReLU activation. MaxPooling2D follows each block to reduce the spatial dimension and incrementally increase the number of feature channels from 32, 64, 128, to 256. The center of architecture (bottleneck) has the smallest resolution of 32x32 with 512 feature channels. This allows the network to capture global contextual information from the input image. After the bottleneck, the decoder path performs upsampling using Conv2DTranspose to restore the spatial resolution to its original size. Each upsampling stage is combined with features from the corresponding encoder through a concatenate operation, helping to retain spatial details lost during downsampling. The uniqueness of this architecture lies in the use of residual connections in each block, which helps overcome the vanishing gradient problem and enables deeper network training. In addition, the use of dropouts in some layers helps prevent overfitting. The total model parameters reach over 7 million, indicating high complexity and capacity to learn important features in segmentation tasks. The architecture ends with a convolutional layer that produces an output the same size as the input (512x512x1), suitable for segmentation tasks requiring pixel-wise prediction. Using BatchNormalization throughout the network helps speed up training and improve model stability.

C. Dense U-Net

The Dense U-Net architecture is a deep neural network used for medical image segmentation, combining U-Net and DenseNet's advantages. Based on Table 3, this architecture begins with an input layer that receives a 512x512x1 resolution image. An initial convolution layer processes this image, generating 32 features while maintaining spatial resolution, followed by batch normalization and activation layers. The concept of "concatenate"

Table 2: Res U-Net network architecture.

Layer	Output Shape	Output Shape	Parameter	Functions
Input	(512, 512, 1)	0	Input layer	Input
Conv2D + BN + ReLU	(512, 512, 32)	448	First convolution layer	Conv2D + BN + ReLU
MaxPool2D	(256, 256, 32)	0	Downsampling	MaxPool2D
Conv2D + BN + ReLU	(256, 256, 64)	18,496	Residual block	Conv2D + BN + ReLU
MaxPool2D	(128, 128, 64)	0	Downsampling	MaxPool2D
Conv2D + BN + ReLU	(128, 128, 128)	73,856	Residual block	Conv2D + BN + ReLU
MaxPool2D	(64, 64, 128)	0	Downsampling	MaxPool2D
Conv2D + BN + ReLU	(64, 64, 256)	295,168	Residual block	Conv2D + BN + ReLU
MaxPool2D	(32, 32, 256)	0	Downsampling	MaxPool2D
Conv2D + BN + ReLU	(32, 32, 512)	1,180,160	Bridge/bottleneck	Conv2D + BN + ReLU
Conv2DTranspose	(64, 64, 256)	524,544	Upsampling	Conv2DTranspose
Conv2D + BN + ReLU	(64, 64, 256)	590,080	Decoder block	Conv2D + BN + ReLU
Conv2DTranspose	(128, 128, 128)	131,200	Upsampling	Conv2DTranspose
Conv2D + BN + ReLU	(128, 128, 128)	147,584	Decoder block	Conv2D + BN + ReLU
Conv2DTranspose	(256, 256, 64)	32,832	Upsampling	Conv2DTranspose
Conv2D + BN + ReLU	(256, 256, 64)	36,928	Decoder block	Conv2D + BN + ReLU
Conv2DTranspose	(512, 512, 32)	8,224	Upsampling	Conv2DTranspose
Conv2D + BN + ReLU	(512, 512, 32)	9,248	Decoder block	Conv2D + BN + ReLU
Conv2D	(512, 512, 1)	33	Output layer	Conv2D

is applied to combine features from the previous layer with the current layer's output, which improves the network's ability to capture contextual information and details at multiple levels. This process is repeated with an increasing depth of features in each successive layer. Each transition layer (max pooling) reduces the spatial dimension while increasing the channel depth, and a dropout layer is applied to reduce overfitting. This sequence continues until the final layer, producing the required segmentation output. This architecture emphasizes dense connections between layers to maintain the flow of gradients and information during training, which is crucial for achieving precise segmentation.

3.2. Model training and parameters

In the process of training the model to improve the accuracy of image segmentation, two main techniques are applied, namely data augmentation and early stopping mechanism. Data augmentation is an important technique to enrich the dataset by creating variations from the existing data. This helps the model become more robust to input variations, reduces overfitting, and improves generalization ability to new data. In the implementation, various transformations are applied, such as rotation to rotate the image by a certain angle, translation to shift the image horizontally and vertically, cropping to take a part of the image, zoom to enlarge or reduce a part of the image, horizontal flipping to reflect the image horizontally and shear to change the perspective of the image. Pada eksperimen untuk fragmentasi kumpulan datanya dibagi secara eksplisit yaitu training set 70%, validasi set 15% dan testing 15%.

This augmentation process is implemented using Keras' ImageDataGenerator, which allows augmentation to be done in real-time during training, saves memory as there is no need to save augmentation results, and ensures the same transformation is applied to images and masks consistently. Combined_generator plays an important role in generating batch data in real-time, keeping image and mask pairs aligned, and ensuring consistent transformations between input and target.

An early stopping mechanism is applied to optimize the training process. This mechanism continuously monitors validation loss and stops training if there is no improvement, thus preventing overfitting. The early stopping

Table 3: Dense U-Net network architecture.

Encoder Path Section	Layer	Output Shape	Parameter
Input	Input Layer	(512,512,1)	0
Block 1	Conv2D + BN + ReLU	(512,512,32)	320
	Conv2D + BN	(512,512,32)	9,248
	MaxPool2D	(256,256,32)	0
Block 2	Conv2D + BN + ReLU	(256,256,64)	18,496
	Conv2D + BN	(256,256,64)	36,928
	MaxPool2D	(128,128,64)	0
Block 3	Conv2D + BN + ReLU	(128,128,128)	73,856
	Conv2D + BN	(128,128,128)	147,584
	MaxPool2D	(64,64,128)	0
Block 4	Conv2D + BN + ReLU	(64,64,256)	295,168
	Conv2D + BN	(64,64,256)	590,080
	MaxPool2D	(32,32,256)	0
Bottleneck			
Bridge	Conv2D + BN + ReLU	(32,32,512)	1,180,160
	Conv2D + BN	(32,32,512)	2,359,808
Decoder Path			
Block 5	UpConv2D	(64,64,256)	524,544
	Conv2D + BN + ReLU	(64,64,256)	590,080
Block 6	UpConv2D	(128,128,128)	131,200
	Conv2D + BN + ReLU	(128,128,128)	147,584
Block 7	UpConv2D	(256,256,64)	32,832
	Conv2D + BN + ReLU	(256,256,64)	36,928
Block 8	UpConv2D	(512,512,32)	8,224
	Conv2D + BN + ReLU	(512,512,32)	9,248
Output	Conv2D	(512,512,1)	33

parameter is set with a patience of 10 epochs, which means the system will wait for 10 epochs before stopping the training if there is no performance improvement. The training configuration is set with a maximum of 100 epochs, 50 steps per epoch, and 25 validation steps, with a batch size adjusted to the needs and memory capacity.

This approach provides various benefits regarding resource efficiency, such as saving computation time, optimizing GPU/CPU usage, and avoiding unnecessary training. In terms of model quality, combining these techniques helps prevent overfitting, improve generalization, and produce more robust models. Best practices include continuous monitoring of validation metrics, use of callbacks for automation, maintaining a balance between augmentation and data authenticity, and consistency of input-output transformation. The overall approach reflects the combination of modern deep learning techniques with proven practices in computer vision model development, creating an efficient and effective training framework to improve the performance of image segmentation models.

The final model comes with more than 7 million learned parameters, reflecting the complexity and depth of the architecture that has been developed. The training process involves iterative weight adjustment based on gradient descent optimization with backpropagation, performed on large datasets to improve the model's ability to produce accurate and reliable predictions. This architecture summary demonstrates a systematic and innovative approach to addressing the challenges of advanced image analysis. With the depth and breadth of the model, the architecture

Table 4: Number of model architecture parameters.

Architectural Methods	Trainable Parameters	Non-trainable Parameters	Total Parameters
U-Net	7,765,409	5,888	7,771,297
Res-UNet	8,115,073	5,888	8,120,961
Dense-UNet	101,514,882	11,776	101,526,658

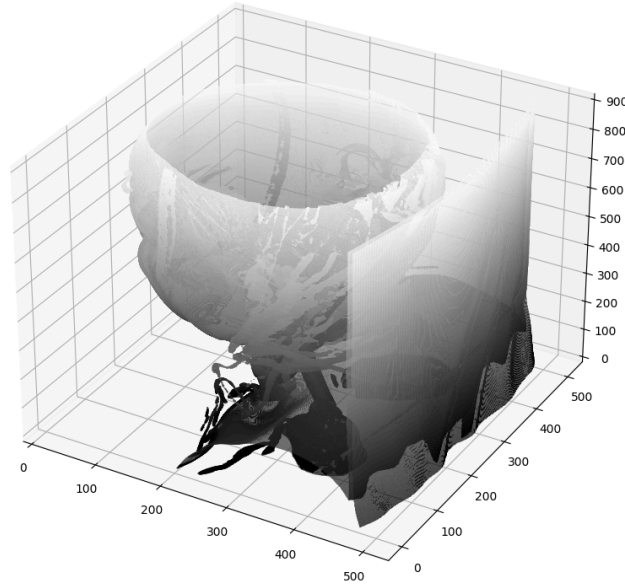


Fig. 2: 3D reconstruction from DICOM data.

promises significant improvements in performance for applications that require advanced image analysis, such as object recognition, segmentation, and feature detection.

Table 4 compares the number of trainable and untrainable parameters in three model architectures: U-Net, Res-U-Net, and Dense-U-Net. From the data presented, the U-Net model has 7771297 parameters, of which 7765409 are trainable and 5888 are untrainable. Meanwhile, the Res-U-Net model has slightly more parameters, with 8120961 parameters, of which 8115073 are trainable, and the remaining 5888 are untrainable. In the Dense-U-Net model, there is a significant spike in the number of parameters, reaching 101526658 parameters, with 101514882 trainable and 11776 untrainable parameters. This analysis indicates significant variations in complexity and learning capacity between the models, Dense-U-Net being the most complex model based on the total number of parameters.

3.3. 3D Reconstruction

The segmentation results are then integrated into the 3D reconstruction and visualization process. Algorithms such as Marching Cubes are used to create 3D surface models from the segmentation output, which enables visualization of the jawbone in three dimensions, as seen in Fig. 2. This visualization is essential for clinical analysis and evaluation of the segmentation performed by U-Net. With the resulting 3D visualization, a doctor or surgeon can view and assess the jawbone model from various angles, which helps plan medical procedures, such as jaw surgery. The clinical applications of this U-Net segmentation cover a wide range of aspects, from diagnosis and surgical planning to surgical simulation. Accurate segmentation of the mandible and maxilla allows clinicians to identify problem areas better, plan surgical paths, and reduce the risk of errors during medical procedures. The integration of U-Net segmentation with 3D reconstruction and visualization in this study is a precision medical analysis tool and a valuable clinical decision support system in jaw health and surgery.

4. Results and Discussion

This chapter discusses two main aspects: 3D visualization and reconstruction results and analysis of model performance in segmentation. The first section describes the visualization quality and details of the 3D reconstruction produced by the segmentation process. The second section analyzes the segmentation results using evaluation metrics and discusses the advantages and limitations of the applied methods.

4.1. Pre-processing

In this process, the study used the Nibabel library to read and process NIfTI image files. This library allows the transformation of 3D volume data into a format ready for further analysis. The file reading process begins using the `load_nifti_file()` function, which extracts image and segmentation data from input files stored in NIfTI format.

The next stage of pre-processing involves a series of complex data transformations. Data normalization is performed by dividing all pixel intensity values by the maximum value, resulting in a range of values between 0 and 1. This aims to standardize the intensity scale of the image, which is essential to ensure consistency in subsequent analysis and modeling. In addition to normalization, channel dimensions, and data matrix are added and transformed to prepare a data structure compatible with further processing needs, especially in deep learning and medical image analysis.

The data transformation involves using the `prepare_slices()` function, which converts a 3D volume into a series of 2D slices with appropriate dimensions. In this study, the image and segment data were transformed from their original form into a (250, 512, 512, 1) format, which allowed each slice to be processed independently. This approach facilitates in-depth analysis of each image slice while retaining the complex structural information of the original mandibular volume data.

4.2. 3D Visualization and Reconstruction

Fig. 3 presents the results of medical image segmentation, showing five different layers: original image, manual segmentation (ground truth), segmentation predicted by the U-Net model, segmentation predicted by the Res U-Net model, and segmentation predicted by the Dense U-Net model. The first column contains input or slices of medical images, most likely obtained from imaging techniques such as CT or MRI. This image displays the internal structure of the body, such as a specific organ or tissue, which is the segmentation target. The second column shows the ground truth the manual segmentation performed by medical experts. This ground truth serves as a reference standard or actual data used to evaluate the performance of the automatic segmentation model.

The ground truth precisely marks important areas such as organ boundaries, abnormal tissues, or tumors. Any difference between the model's prediction and the ground truth may indicate a prediction error that could affect the accuracy of diagnosis and clinical decision-making. The third, fourth, and fifth columns display the segmentations predicted by the automated model. In each instance, these predictions are compared with the ground truth to assess how each model successfully replicates the manual segmentation. Visually, some predictions show high accuracy, while others show discrepancies, such as imperfect shapes, blurred boundaries, or areas that are not segmented correctly. This highlights the challenges of automated segmentation, especially when dealing with complex and highly variable objects or networks.

Visualizations such as these are important to ensure that the model segmentation aligns with the ground truth or manual annotations performed by medical experts. As such, this overlay is a tool to evaluate the model's accuracy, validate the accuracy of predictions, and identify potential segmentation errors, such as undetected areas or imprecise boundaries.

4.3. Analysis of Results

The comparative evaluation conducted on the three deep learning architectures in medical image segmentation has yielded significant findings in Table 5. This study compares the performance of U-Net, Res-Unet, and Dense-Unet through a comprehensive set of measurements covering five key parameters. The analysis results show the dominance of the U-Net architecture in most of the evaluation metrics used. In terms of the Dice Similarity

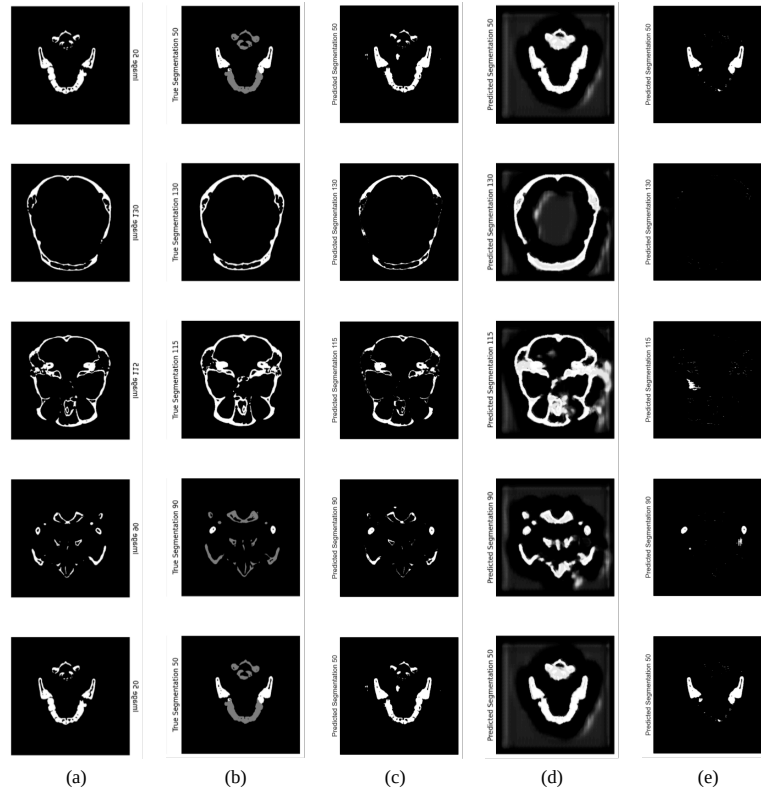


Fig. 3: Segmentation results using the model: (a) Original Image, (b) Ground Truth, (c) U-Net model, (d) Res U-Net Model, (e) Dense U-Net.

Table 5: Quantitative comparison of U-Net, Res U-Net and Dense U-Net.

Architectural Methods	DSC (%)	Precision (%)	Sensitivity (%)	Specificity (%)	Training Time Cost (minutes)
U-Net	92.96	95.90	93.22	99.74	755
Res-Unet	87.27	95.45	92.32	99.71	785
Dense-Unet	80.23	90.36	88.37	80.12	767

Coefficient (DSC), U-Net recorded a very satisfactory performance of 92.96%, substantially outperforming the results achieved by Res-Unet (87.27%) and Dense-Unet (80.23%). U-Net's superiority is also manifested in the Precision parameter, where this architecture achieves an accuracy rate of 95.90%, while Res-Unet and Dense-Unet register 95.45% and 90.36%, respectively.

Regarding sensitivity, the U-Net architecture again demonstrated superiority with a true positive detection rate of 93.22%. Res-Unet followed this performance with 92.32% and Dense-Unet with 88.37%. The specificity parameter confirms U-Net's dominance with 99.74%, surpassing the capability of Res-Unet (99.71%) and Dense-Unet (80.12%) in identifying true negatives. Interestingly, the evaluation of the temporal aspect shows a different pattern. Dense-Unet showed relatively better computational efficiency with a processing time of 767 minutes, compared to U-Net, which required 755 minutes, and Res-Unet, which required 785 minutes. This observation indicates a trade-off between accuracy and computational efficiency that needs to be considered in practical implementation.

The comprehensive evaluation results show that the U-Net architecture performs superior in most evaluation parameters. However, practical considerations regarding the balance between accuracy and computational efficiency remain crucial in selecting the optimal architecture for implementation in specific contexts.

The performance analysis of the U-Net model shows superior results based on three key metrics in Fig. 4, proving its superiority over Res-UNet and Dense-UNet. In the Model Loss graph (Fig. 4a), U-Net achieves the fastest convergence, decreasing the loss from 0.7 to 0.1 within the first 20 epochs. The training and validation loss curves show optimal parallelism and reach stability at 0.02 after the 20th epoch, proving the best generalization

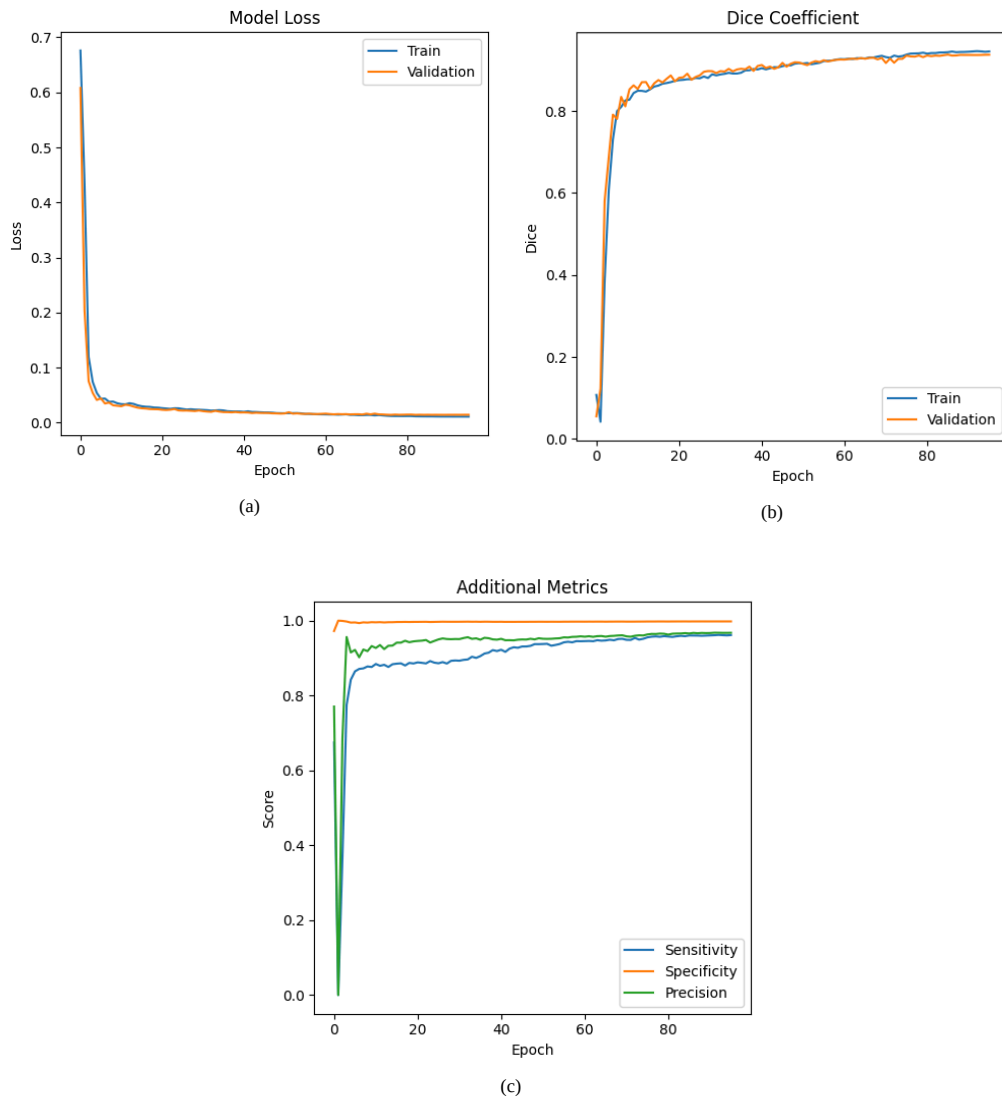


Fig. 4: Performance Analysis of U-Net Model Loss (a), Dice Coefficient (b) and Additional Metrics (c) in Training and Validation Processes.

capability without overfitting. The segmentation performance measured by the Dice Coefficient (Fig. 4b) shows the highest efficiency with a drastic increase from 0 to 0.8 in the first 20 epochs, reaching an optimal value of 0.93 in the 100th epoch, surpassing the performance of Res-UNet (0.85-0.9) and Dense-UNet (0.9-0.95 with instability). The perfect parallelism between the training and validation dice coefficient curves proves the superior stability in segmentation. Additional Metrics (Fig. 4c) confirmed the dominance of U-Net, with sensitivity, specificity, and precision achieving the fastest improvement in the first 20 epochs and consistent values above 0.9 after the 40th epoch. U-Net proved its absolute superiority with the fastest convergence in the first 20 epochs, the best training-validation stability, and the highest segmentation accuracy (dice coefficient 0.93). It outperformed Res-UNet and Dense-UNet in all aspects of measurement, including learning speed, stability, and final accuracy.

The performance analysis of the Res-UNet model shows a good learning pattern but not as optimal as U-Net, based on the three metrics in Fig. 5. In the Res-UNet Loss graph (Fig. 5a), the model experiences a relatively slow decrease in loss from 0.6 to 0.1 in the first 20 epochs, followed by convergence to a value of 0.02-0.03 after the 40th epoch. The training and validation loss curves show a larger gap than U-Net, although still better than Dense-UNet. The Dice Coefficient evaluation (Fig. 5b) shows a slower performance improvement than U-Net, starting from a value of 0.2-0.3 and increasing gradually until it reaches a plateau of 0.85-0.9 after the 80th epoch. The difference between the training and validation dice coefficient (0.02-0.03) shows good generalization, although not as good as U-Net. Additional Metrics (Fig. 5c) display moderate performance, with Specificity and Precision reaching values

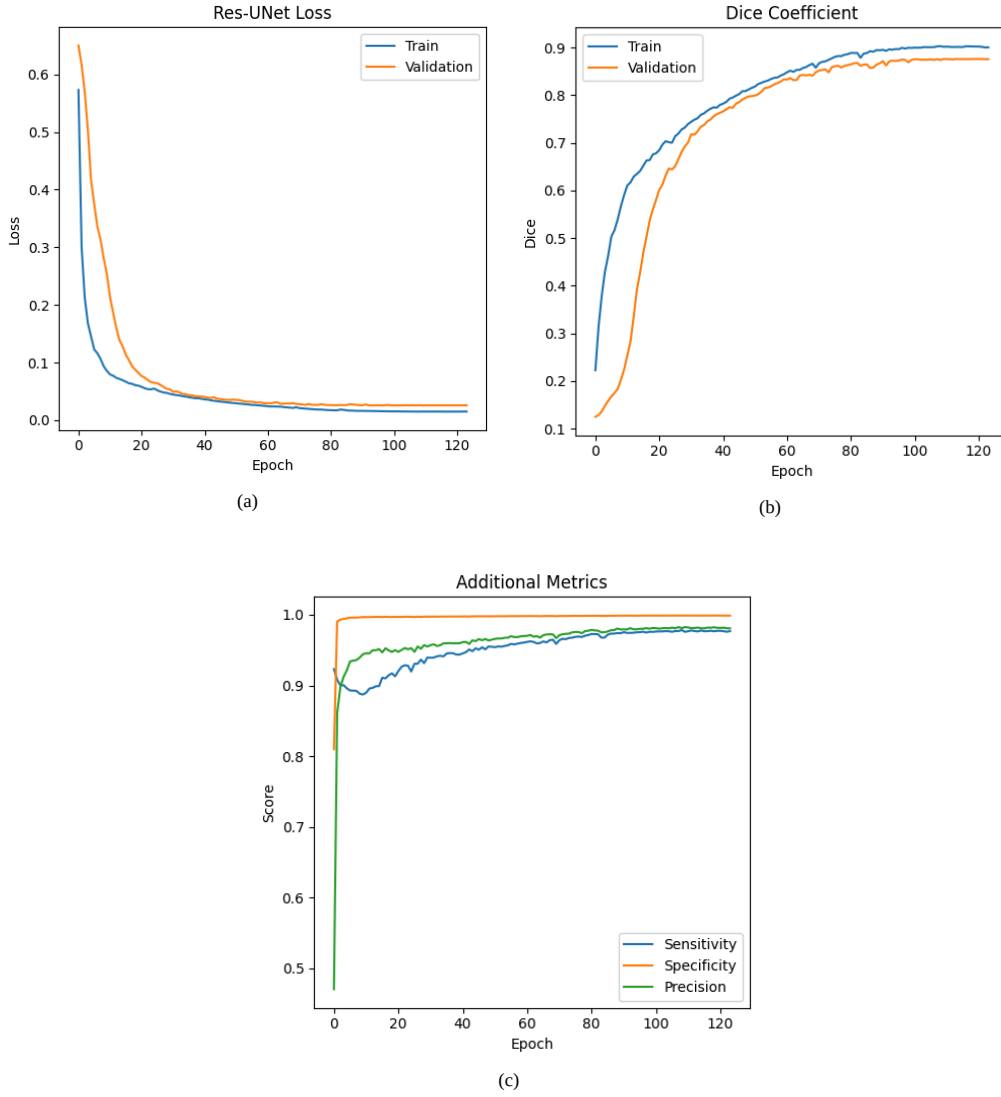


Fig. 5: Performance Analysis of Res U-Net Model Loss (a), Dice Coefficient (b) and Additional Metrics (c) in Training and Validation Processes.

close to 1.0, while Sensitivity reaches 0.95, indicating good accuracy in pixel classification, although not as high as U-Net. Overall, Res-UNet demonstrated intermediate performance among the three models, with convergence and generalization slower than U-Net but better than Dense-UNet. This is evidenced by a low combination loss (0.02-0.03), a good dice coefficient (0.85-0.9), and an additional metric that reaches the optimal value (≥ 0.95), although it requires more epochs to reach this value than U-Net.

The performance analysis of Dense-UNet based on Fig. 6 shows less efficient learning characteristics in three aspects of measurement. In the Dense-UNet Loss graph (Fig. 6a), the model shows instability with a significant gap between training and validation loss in the first 10 epochs, where the training loss reaches 0.05, but the validation loss decreases very slowly. The model takes longer to reach stability until the 20th epoch with a loss value of 0.02-0.03, indicating an inefficient learning process. The Dice Coefficient evaluation (Fig. 6b) shows a clear performance imbalance, where the training curve rapidly increases to 0.8 within the first 10 epochs. In contrast, the validation curve lags far behind and increases very slowly. Convergence is only achieved after the 40th epoch at 0.9-0.95, indicating an inefficient learning process. Additional Metrics (Fig. 6c) show variations in performance with Specificity reaching 1.0, but Precision and Sensitivity show instability before finally reaching 0.95-0.97. Overall, Dense-UNet demonstrated lower performance than U-Net and Res-UNet, characterized by slower learning, a larger training-validation gap, and instability in evaluation metrics. It requires more epochs to achieve convergence and shows lower learning efficiency than the previous architecture.

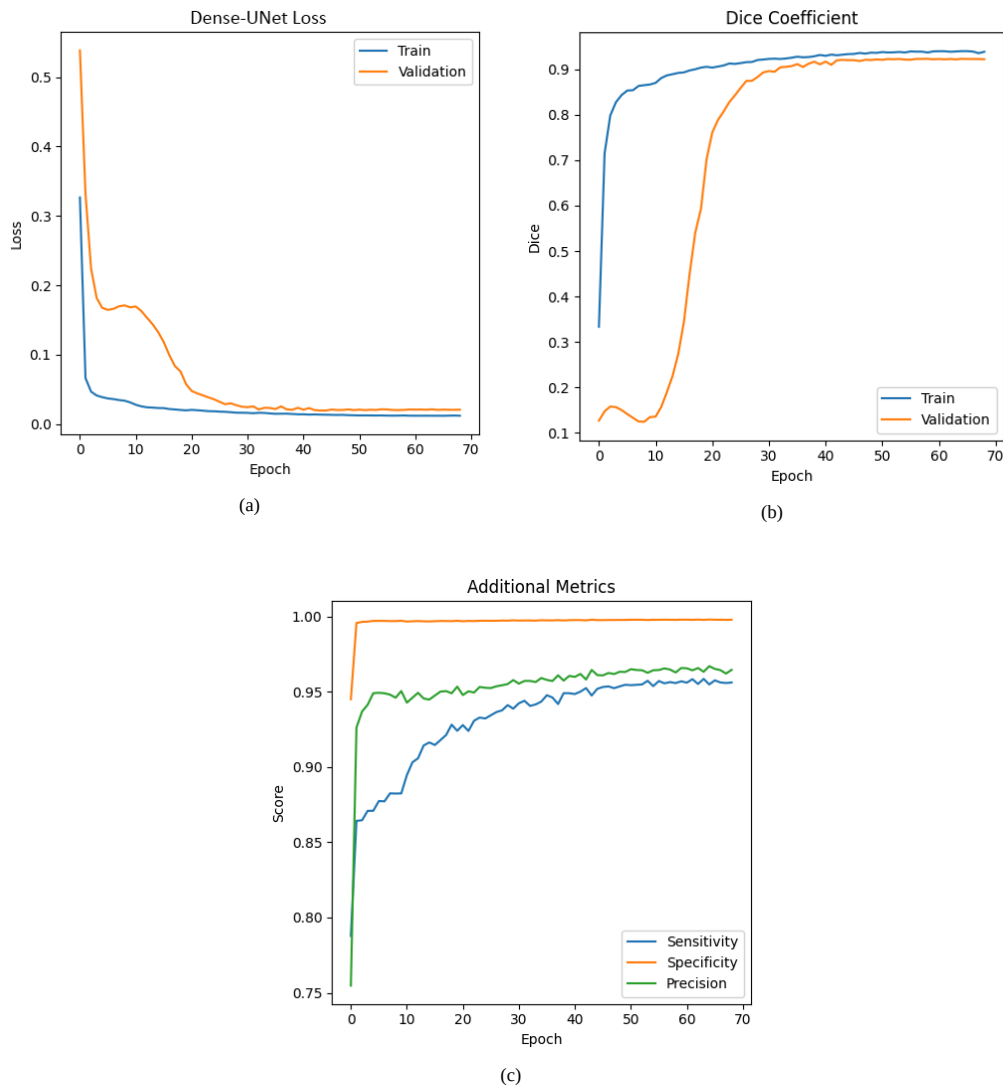


Fig. 6: Performance Analysis of Dense U-Net Model Loss (a), Dice Coefficient (b) and Additional Metrics (c) in Training and Validation Processes.

5. Conclusion

Based on the evaluation conducted on three deep learning architectures (U-Net, Res-U-Net, and Dense-U-Net), the results show that U-Net dominates most of the evaluation parameters. In terms of the Dice Similarity Coefficient (DSC), U-Net achieves the highest score of 92.96%, significantly outperforming Res-U-Net (87.27%) and Dense-U-Net (80.23%). U-Net's superiority is also evident in the precision parameter (95.90%) and sensitivity (93.22%), which are better than those of Res-U-Net and Dense-U-Net. Additionally, U-Net also excels in specificity (99.74%), surpassing Res-U-Net (99.71%) and Dense-U-Net (80.12%). U-Net's superiority in DSC demonstrates its better capability to capture the similarity between the segmentation results and the ground truth, which is crucial in medical segmentation applications. A higher DSC indicates that U-Net is more stable in handling variations in the shape and size of objects in medical images, making it more reliable in clinical environments. The implications of this study are highly significant in the development of medical image segmentation methods, particularly in improving the accuracy of anatomical structure detection in medical imaging. These results reaffirm that the selection of deep learning models should consider dataset characteristics and the specific objectives of clinical applications. Furthermore, this study is a foundation for developing hybrid models or modifications of the U-Net architecture to enhance computational efficiency and segmentation performance under various imaging conditions. Future research may optimize the architecture while considering computational time efficiency and exploring transfer learning and data augmentation techniques to improve model performance across various medical application domains.

CRedit Authorship Contribution Statement

Mambaul Izzi: Conceptualization, Methodology, Software, Investigation, Resources, Data Curation, Writing – Original Draft, Visualization, Funding Acquisition. **Chastine Fatichah:** Validation, Formal analysis, Resources, Writing – Review & Editing, Supervision, Project Administration. **Hadziq Fabroyir:** Writing – Review & Editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Declaration of Generative AI and AI-assisted Technologies in The Writing Process

The authors used generative AI to improve the writing clarity of this paper. They reviewed and edited the AI-assisted content and take full responsibility for the final publication.

References

- [1] Z. Huang, T. Xia, J. Kim, L. Zhang, and B. Li, "Combining CNN with Pathological Information for the Detection of Transmissive Lesions of Jawbones from CBCT Images," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 2972–2975, 2021, doi: 10.1109/EMBC46164.2021.9630692.
- [2] I. B. Păvăloiu, A. Vasilăţeanu, N. Goga, I. Marin, C. Ilie, A. Ungar, and I. Pătraşcu, "3D dental reconstruction from CBCT data," *2014 Int. Symp. Fundam. Electr. Eng. ISFEE 2014*, 2015, doi: 10.1109/ISFEE.2014.7050617.
- [3] S. Saluja, M. C. Trivedi, and A. K. Dubey, "U-Net Variants for Brain Tumor Segmentation: Performance and Limitations," *2023 Int. Conf. Comput. Intell. Commun. Technol. Networking, CICTN 2023*, pp. 612–616, 2023, doi: 10.1109/CICTN57981.2023.10141477.
- [4] P. U. Augmented, "Nextmed: Automatic Imaging Segmentation, 3D Reconstruction, and 3D Model Visualization Platform Using Augmented and Virtual Reality," 2020, doi: 10.3390/s20102962.
- [5] T. Selzner et al., "3D U-Net Segmentation Improves Root System Reconstruction from 3D MRI Images in Automated and Manual Virtual Reality Work Flows," *Plant Phenomics*, vol. 5, p. 76, 2023, doi: 10.34133/plantphenomics.0076.
- [6] M. L. Allaoui, M. S. Allili, and A. Belaid, "HA-U3Net: A modality-agnostic framework for 3D medical image segmentation using nested V-Net structure and hybrid attention," *Knowledge-Based Systems*, vol. 327, p. 114127, 2025, doi: 10.1016/j.knosys.2025.114127.
- [7] W. F. Chen, H. Y. Ou, K. H. Liu, Z. Y. Li, C. C. Liao, S. Y. Wang, W. Huang, Y. F. Cheng, and C. T. Pan, "In-series u-net network to 3d tumor image reconstruction for liver hepatocellular carcinoma recognition," *Diagnostics*, vol. 11, no. 1, 2021, doi: 10.3390/diagnostics11010011.
- [8] V. S. Nguyen, M. H. Tran, and H. M. Quang Vu, "An Improved Method for Building A 3D Model from 2D DICOM," *Proc. - 2018 Int. Conf. Adv. Comput. Appl. ACOMP 2018*, pp. 125–131, 2018, doi: 10.1109/ACOMP.2018.00027.
- [9] K. Chrz, J. Bruthans, J. Ptáček, and Č. Štuka, "A cost-affordable methodology of 3D printing of bone fractures using DICOM files in traumatology," *Journal of Medical Systems*, vol. 48, no. 1, p. 66, 2024, doi: 10.1007/s10916-024-02084-w.
- [10] A. Sha, E. R. Adwaith Krishna, D. S. Menon, and T. Anjali, "Enhancing Segmentation Efficiency: A 2D U-Net Approach with 3D-to-2D Conversion for Medical Image Analysis," *7th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2023 - Proc.*, no. Iceca, pp. 206–211, 2023, doi: 10.1109/ICECA58529.2023.10395625.
- [11] A. Gamal, K. Bedda, N. Ashraf, S. Ayman, M. Abdallah, and M. A. Rushdi, "Brain Tumor Segmentation using 3D U-Net with Hyperparameter Optimization," *2021 3rd Nov. Intell. Lead. Emerg. Sci. Conf.*, pp. 269–272, 2021, doi: 10.1109/NILES53778.2021.9600556.
- [12] L. Melerowitz et al., "Design and evaluation of a deep learning-based automatic segmentation of maxillary and mandibular substructures using a 3D U-Net," *Clinical and Translational Radiation Oncology*, vol. 47, p. 100780, 2024, doi: 10.1016/j.ctro.2024.100780.
- [13] E. K. Rutoh, Q. Zhi Guang, N. Bahadar, R. Raza, and M. S. Hanif, "GAIR-U-Net: 3D guided attention inception residual u-net for brain tumor segmentation using multimodal MRI images," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 6, p. 102086, 2024, doi: 10.1016/j.jksuci.2024.102086.
- [14] H. Mzoughi, I. Njeh, M. Ben Slima, and A. Benhamida, "Glioblastomas brain Tumor Segmentation using Optimized U-Net based on Deep Fully Convolutional Networks (D-FCNs)," *2020 Int. Conf. Adv. Technol. Signal Image Process. ATSIP 2020*, pp. 1–6, 2020, doi: 10.1109/ATSIP49331.2020.9231681.
- [15] S. Chihati and D. Gaceb, "A review of recent progress in deep learning-based methods for MRI brain tumor segmentation," *2020 11th Int. Conf. Inf. Commun. Syst. ICICS 2020*, pp. 149–154, 2020, doi: 10.1109/ICICS49469.2020.239550.
- [16] B. Jiang et al., "Deep learning for brain tumor segmentation in multimodal MRI images: A review of methods and advances," *Image and Vision Computing*, vol. 156, p. 105463, 2025, doi: 10.1016/j.imavis.2025.105463.
- [17] Z. Zhou, Y. Chen, A. He, X. Que, K. Wang, R. Yao, and T. Li, "NKUT: Dataset and Benchmark for Pediatric Mandibular Wisdom Teeth Segmentation," *IEEE J. Biomed. Heal. Informatics*, vol. 28, no. 6, pp. 3523–3533, 2024, doi: 10.1109/JBHI.2024.3383222.