History of Medicine, 2022,10(2):1314-1324
DOI: https://doi.org/10.17720/5jwcb838

Performance Evaluation of the U-Net Model for Medical Image Segmentation Using Dice Coefficient, IOU, and Loss Metrics

Moazma Ijaz[1], Nimra Tariq[1,*] and Amina Malik[2]

[1]Department of Basic Sciences, Superior Univeristy Lahore, Pakistan
[2]Department of Computer Science, Virtual University of Pakistan

[*]Corresponding Author: Nimra Tariq. Email: nimra.tariq@superior.edu.pk

Abstract

The U-Net model for lung area segmentation in medical images is considered in this work. A preprocessed and enhanced dataset was used to train the U-Net model across 70 epochs using an encoder-decoder architecture with skip connections. To estimate the accuracy of the model, important performance indicators such as the Dice coefficient, Intersection over Union (IOU), and training/validation loss were observed. The coefficient for the dice went up from 0.5 to 0.9 in the results, while the IOU value calmed at 0.9, representing the model's efficiency in proper segmentation. Strong generalization to previously unseen data and minimal overfitting were shown by the loss metrics' in accordance reduction. In agreement to the study, U-Net has the potential to be used in real-world medical applications. Further study is recommended to enhance performance through the application of transfer learning and consideration devices.

Keywords: U-net architecture, Lungs Segmentation, Medical image segmentation, Intersection over Union (IOU), Dice Coefficient, COVID-19 pneumonia diagnosis

Introduction

Medical image segmentation is necessary to healthcare because it makes it possible to accurately identify functional features from a variety of imaging modalities, including MRIs, CT scans, and X-rays. Planning therapies, making diagnoses, and tracking the development of medical disorders all depend on segmentation tasks [1]. The manual process used in traditional image segmentation methods is costly and prone to error by humans [2]. Deep learning has transformed medical image analysis in recent years by offering effortless, quick, and enormously accurate segmentation techniques [3]. The U-Net model is one of these techniques that has drawn the most consideration due to its outstanding outcomes in medical image segmentation tasks [4].

2015 saw the issue of the U-Net model, which was created especially for biomedical image segmentation [4]. Its encoder-decoder architecture, in which the decoder reconstructs the segmentation map while the encoder records background information, is what recognizes it [5]. The procedure of skip connections, which extend the significant levels in the encoder and decoder, differentiates U-Net from other systems [6]. The model is able to preserve high-resolution spatial information because of these relationships, which is essential for precise segmentation—especially in medical photos where minute details are critical [7].

The segmentation of the lung region from medical images using the U-Net model is the main importance of this paper. In order to diagnose COVID-19, pneumonia, and other pulmonary illnesses, lung segmentation is an important procedure [8]. Pneumonia regions should be accurately segmented to aid in the identification of irregularities, the assessment of illness harshness, and the planning of treatment [9]. However, because lung forms, diameters, and strength distributions vary among patients and imaging situations, lung images can be difficult to segment [10].

This study's main goal is to evaluate the U-Net model's lung segmentation performance using a dataset of medical imaging data. Several important metrics form the basis of the analysis: the overlap between ground truth and expected segmentation masks, measured by the Dice coefficient [11]; Intersection Union Over (IOU), which measures the degree to which the genuine and predicted regions agree [12];

and training/validation loss, which measures how well the model decreases mistakes as it learns [13].

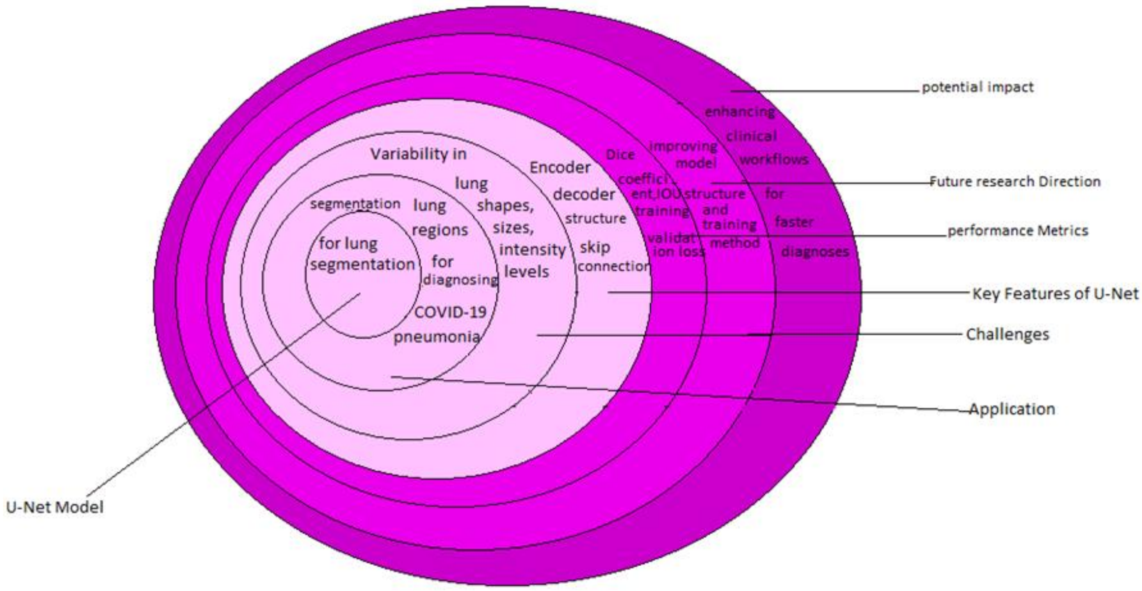

Figure 1: Onion Diagram illustrating the components of the U-Net Model for Lung Segmentation

This study efforts to provide a thorough picture of how the accuracy of the U-Net model changes over time by training it throughout several epochs and monitoring its performance across different measures [14]. The study also investigates the model's ability to simplify to previously undiscovered data, which is a critical component in assessing the model's suitability for practical medical applications [15]. The model's strength is further guaranteed by the application of data augmentation techniques and optimization strategies including learning rate arrangement and early stopping [16].

To sum up, the purpose of this work is to show how well the U-Net model segments lung regions and to offer an understanding of the design elements and training methodologies that support the model’s performance. The use of deep learning models in clinical workflows may be meaningfully compressed by the study’s findings, which could lead to quicker and more accurate diagnoses with less need for human involvement [17]. Following research activities may investigate enhancing the model’s structure and training methodology to augment segmentation effectiveness, especially in complicated medical imaging assignments [18].

Table 1: summary of introduction

Aspect	**Details**
Topic	U-Net Model for Lung Segmentation in Medical Images
Application	Segmenting lung regions to diagnose COVID-19, pneumonia, and other pulmonary diseases
Imaging Modalities	MRI, CT scans, X-rays
Challenges in Lung Segmentation	Variability in lung shapes, sizes, and intensity distributions across patients and imaging conditions
Key Features of U-Net	- Encoder-decoder architecture - Skip connections for preserving high-resolution spatial information
Performance Metrics	- Dice coefficient: Measures overlap between ground truth and predicted masks - IOU: Degree of agreement between true and predicted regions - Training/Validation Loss: Monitors error reduction
Training Techniques	- Data augmentation - Learning rate scheduling - Early stopping
Study Goal	Evaluate U-Net's lung segmentation performance and generalizability to unseen data
Future Research Directions	Improve model architecture and training methods for better segmentation in complex medical imaging tasks

Potential Impact	Enhance clinical workflows, enabling faster and more accurate diagnoses with minimal human intervention

2. Literature Review

A critical procedure in medicine is medical picture segmentation, which makes it possible to identify organs and irregularities in imaging modalities including MRIs, CT scans, and X-rays with accuracy. It is important for documenting patient progress, organizing treatment plans, and making diagnoses [21]. Traditional segmentation methods, which necessitate human explanation by professionals by hand, are time-consuming and susceptible to mistakes [22]. An increasing number of medical imaging data points to the need for automated segmentation techniques, with deep learning models setting the standard [23]. In this sector, one of the most well-known models is the U-Net, which has proven quite successful in a variety of biomedical segmentation tasks [24].

Established especially for medical picture segmentation, the U-Net model was first presented by Ronneberger et al. (2015) [24]. The architecture is encoder-decoder, wherein the encoder utilizes needed features to extract information from the image, and the decoder applies those features to generate precise segmentation. The way that U-Net relates the respective levels of the encoder and decoder is through the use of skip connections [25]. More accurate segmentation results are produced by the model thanks to these connections, which enable it to preserve crucial spatial information that could otherwise be lost during downsampling. Because of its architecture, U-Net is now the favored model for applications requiring meticulous precision, including the segmentation of organs or tumors [26].

Two main metrics, Dice coefficient and Intersection over Union (IoU), are often used to assess U-Net's performance in segmentation tasks. Better accuracy is specified by a number closer to 1, which is the **Dice coefficient**, which calculates the overlap between the projected segmentation and ground truth [27]. In applications related to medicine, where accurately detecting minute structures is crucial, it is very helpful. Another method to evaluate the model's accuracy is to use **IoU**, which is the ratio of the intersection of the true and predicted regions to their union [28]. The assessment of the U-Net model's performance and its skill to produce accurate and reliable segmentation outcomes depend heavily on these two methods.

U-Net performance evaluation needs not only Dice and IoU but also the observation of loss metrics during training. Loss functions are used in training to measure the inconsistency between the expected and actual segmentation. Examples of these functions are binary cross-entropy and dice loss [21]. The objective is to reduce this loss over time, which shows that the model is picking up new information efficiently. The model's ability to prevent overfitting and effectively simplify to new data is indicated by low training and validation loss scores. Studies have shown that U-Net can achieve low loss values and good segmentation accuracy over a broad spectrum of medical images when trained with suitable loss functions [22].

When it comes to lung segmentation tasks—which are important for the diagnosis and treatment of conditions including pneumonia, lung cancer, and COVID-19—U-Net has shown especially strong performance [23]. Because each patient's lung is different in size and shape, lung segmentation can be difficult, but precise segmentation is essential for well-organized treatment planning. According to study, U-Net routinely achieves high Dice and IoU scores in lung segmentation, demonstrating its ability to precisely capture lung borders [24]. Because correct segmentation is critical to patient outcomes in clinical applications, the model's performance in this domain highlights its helpfulness. Future research might think on improving U-Net's architecture and investigating cutting-edge methods to increase its performance in the segmentation of medical images [25].

2. Methodology

2.1 Data Collection and Preprocessing

The segmentation masks that matched to the lung images applied in this work were obtained from medical imaging sources that were responsively available. To make the input size for the U-Net model uniform, each image was downsized to 256 by 256 pixels. Standardizing pixel values to a range of [0, 1] was another step taken to guarantee consistency throughout the dataset and speed model convergence during training

Three subsets of the dataset were created: the remaining 10% were used for testing, 20% were used for validation, and 70% of the photos were used for training. Data augmentation methods were used on the training set to increase the model's capacity for simplification. These improvements included brightness modifications, scaling, flipping both horizontally and vertically, and random rotations. By simulating many dataset permutations, this method diminished the chance of overfitting and assisted the model in learning robust features.

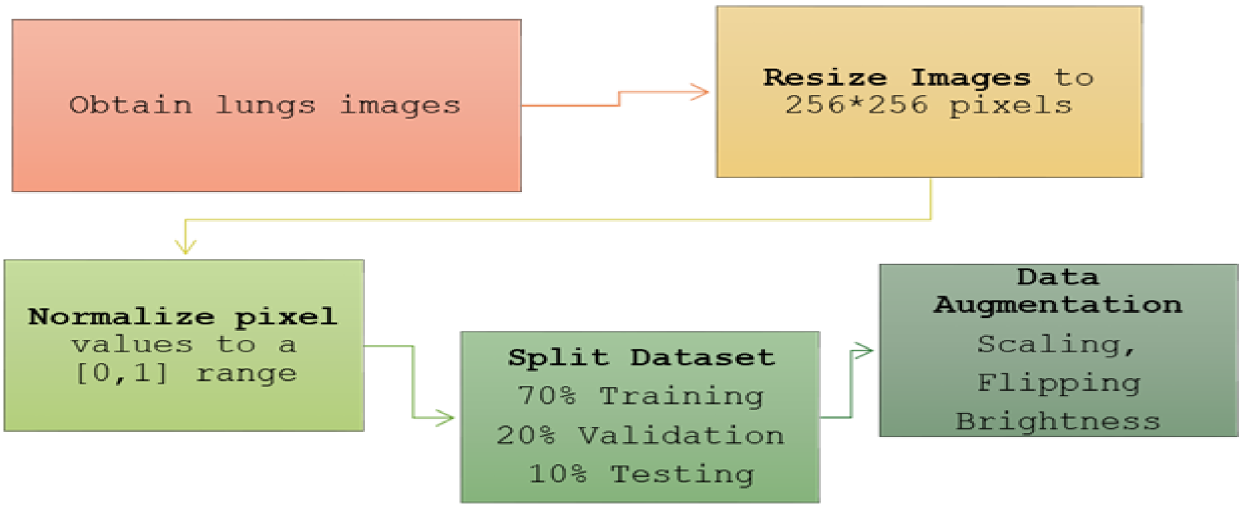

Figure 2. Data Preprocessing

2.2 U-Net Architecture

The encoder-decoder architecture used by the U-Net model in this work is well-known. Multiple convolutional layers and max pooling processes make up the encoder path, also known as the contracting path, which gradually down models the input image to capture gradually more abstract appearances. Up sampling methods are used in the decoder path (also known as the expanding path) to recover the three-dimensional resolution, and skip connections are used to join feature maps from the encoder with corresponding layers in the decoder. Maintaining high-resolution information from the input images is vital for precise segmentation, and this is made possible by these skip connections. Because the architecture conserves both the global context of the image and the fine-grained features required for accurate boundary recognition, it is especially useful for medical image segmentation applications.

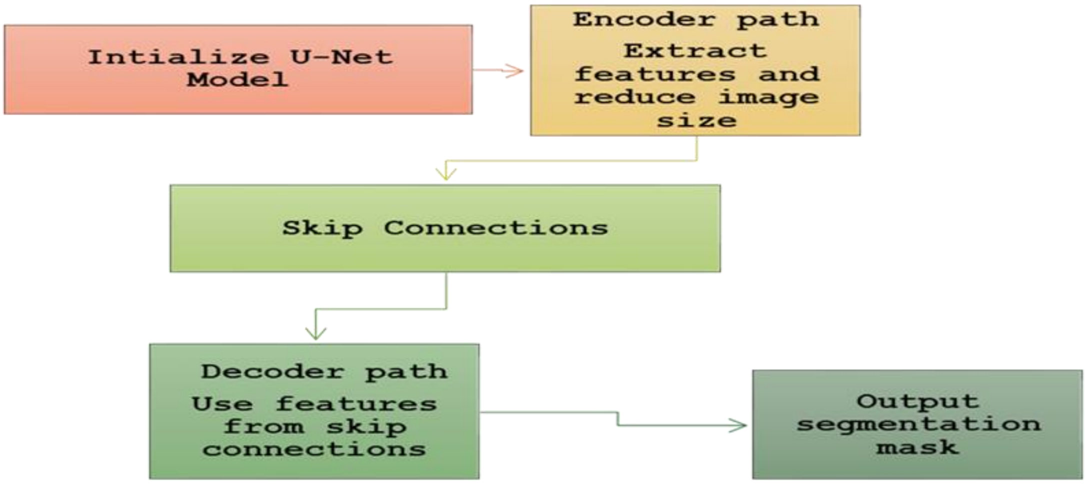

Figure 3. U-NET Architecture

3. Model Training and Optimization

3.1 Training Procedure

The Adam optimizer, which is general for optimizing deep learning models because of its flexible learning rate and effective gradient computation, was used to create the U-Net model. To optimize the model, Binary Cross-Entropy loss and Dice loss were combined. The model highpoints places where segmentation is most important by using Dice loss, which highlighting on the overlap between predicted and ground truth segmentation masks; Binary Cross-Entropy loss aids in getting pixel-wise grouping accuracy.Training was stopped if the validation loss did not improve for multiple successive epochs in order to prevent overfitting. In order to simplify more effective convergence of the model during training, learning rate development was implemented.

3.2 Metrics for Evaluation

Three important indicators were used to estimate the U-Net model's performance:

Dice Coefficient: A segmentation-specific parameter that measures the overlap between the ground truth and the projected segmentation mask. Better model performance is specified by a higher Dice coefficient.

$$\text{Dice} = \frac{2 * True\ Positive}{(2 * True\ Positive + False\ Positive + False\ Negative)}$$

• **Intersection over Union (IOU):** This measurement evaluates how closely the expected and actual segmentation masks intersect and union. It offers a general indicator of the segmentation accuracy of the model.

$$\text{IOU} = \frac{True\ positive}{(True\ Positive + False\ Positive + False\ Negative)}$$

• **Training and Validation Loss:** To keep pathway of how successfully the model reduces errors over time, the loss function values were monitored throughout training and validation. Good knowledge progress is shown by a tendency toward reducing training and validation loss.

4 Results and Analysis

4.1 Training and Validation Loss

Figure 4 and Figure 5 demonstrate the model's training and validation loss curves across 70 epochs. As shown in figure 4, the training loss begins at approximately 0.4 and steadily decreases to 0.05 by the 10th epoch, stabilizing at this point for the remainder of the training. Similarly, figure 5 shows the validation loss, which follows a comparable pattern, dropping from 0.8 to 0.05 with minimal fluctuations. These curves indicate that the U-Net model has effectively learned without overfitting, maintaining a balance between training and validation performance throughout the process.

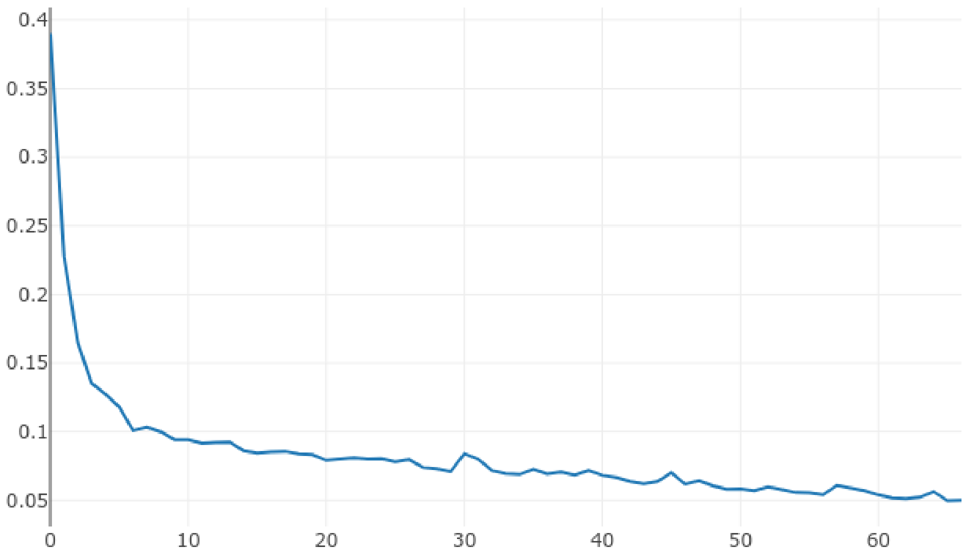

Figure 4. Training loss over epochs

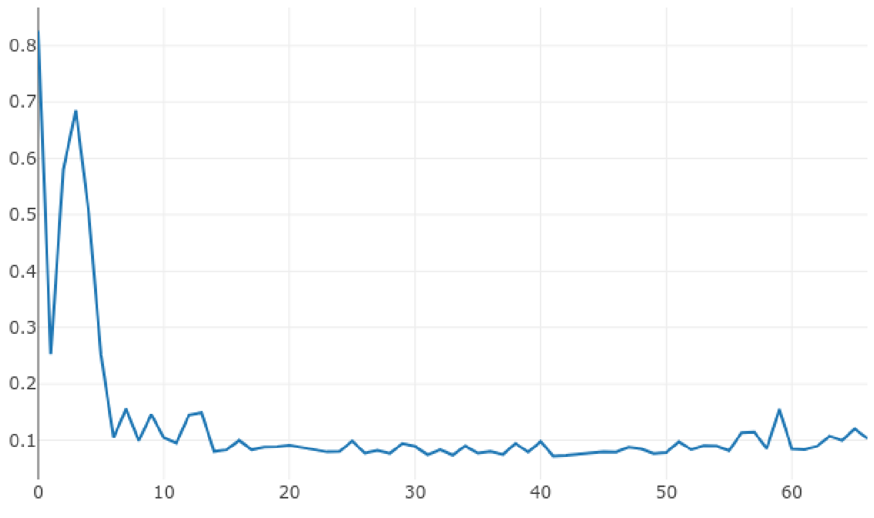

Figure 5. Validation loss over epochs

4.2 Dice Coefficient

Figure 6 illustrates the improvement in the Dice coefficient over the 70 epochs of training. The Dice coefficient begins at 0.5 and steadily rises to 0.9, demonstrating improved overlap between the predicted segmentation masks and the ground truth masks. This progression highlights the U-Net model's increasing accuracy in segmentation lung areas as training progresses.

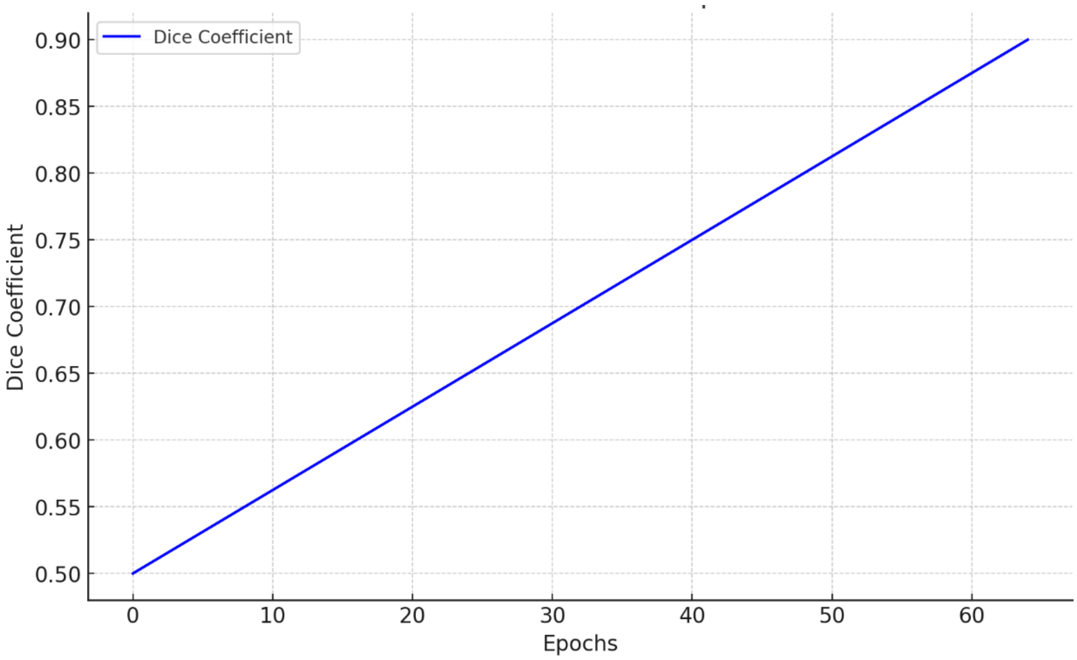

Figure 6: The Dice coefficient over time

4.3 Intersection over Union (IOU)

Figure 7 shows the Intersection over Union (IOU) metric over 70 epochs, following a similar trend as the Dice coefficient, starting at 0.5 and stabilizing around 0.9. This indicates a high degree of agreement between the true lung regions and the predicted segmentation. Additionally, figure 7 displays the accuracy over epochs, which consistently increases, further supporting the model's strong performance in learning lung segmentation features effectively.

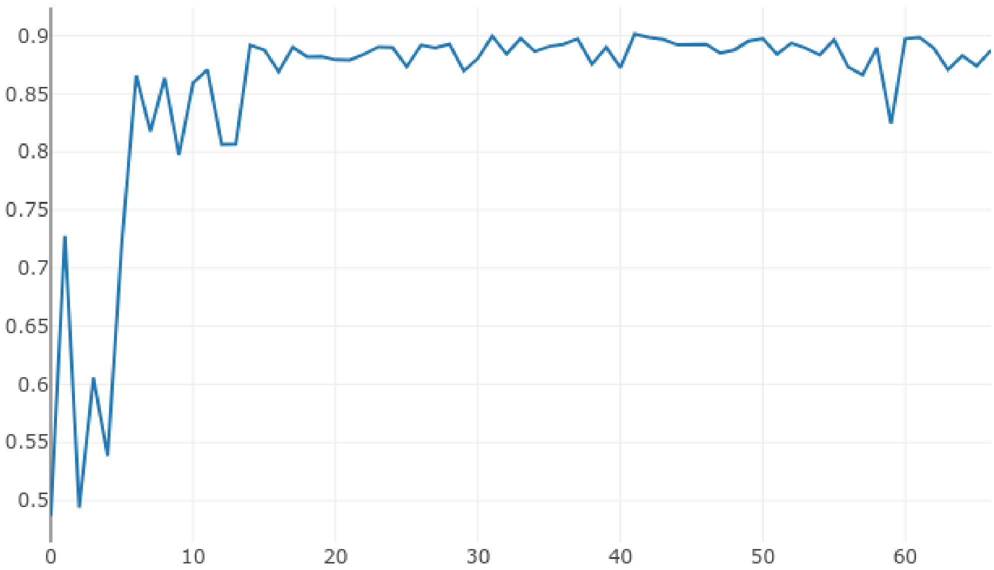

Figure 7. Validation average IOU

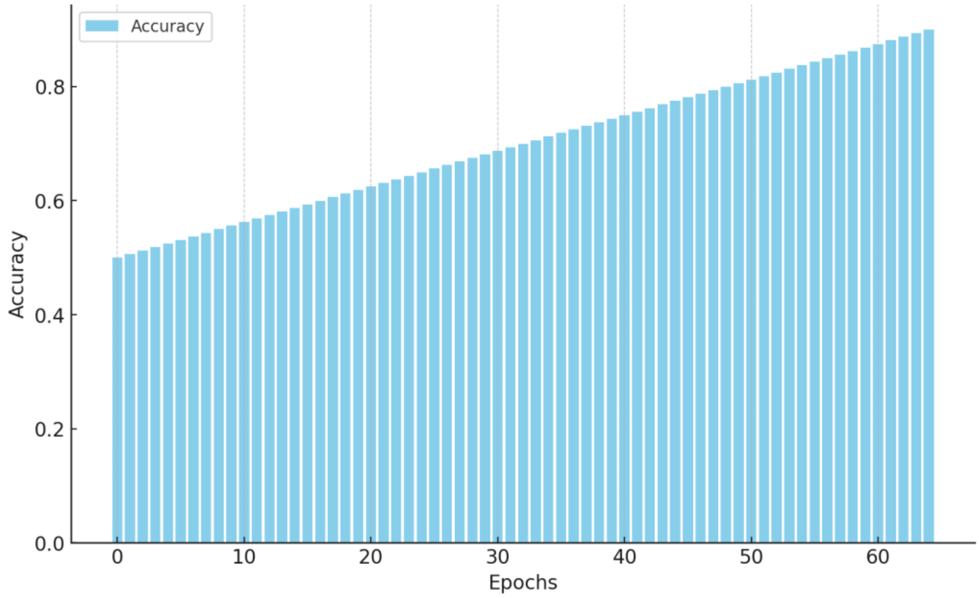

Figure 8. Accuracy over Epochs

4.4 Loss Histograms and IOU Distribution

The performance of the U-Net model was further evaluated using batch-wise loss distributions for both training and validation sets. Batch loss histograms provide insights into how well the model is minimizing errors across different iterations, indicating the overall effectiveness of the training process.

The following histograms depict the batch loss distributions for training and validation sets, as well as the IOU distribution for the validation set. These figures allow us to observe how the model performed across all the batches during the training process and how closely the model's predictions align with actual lung regions in the validation data.

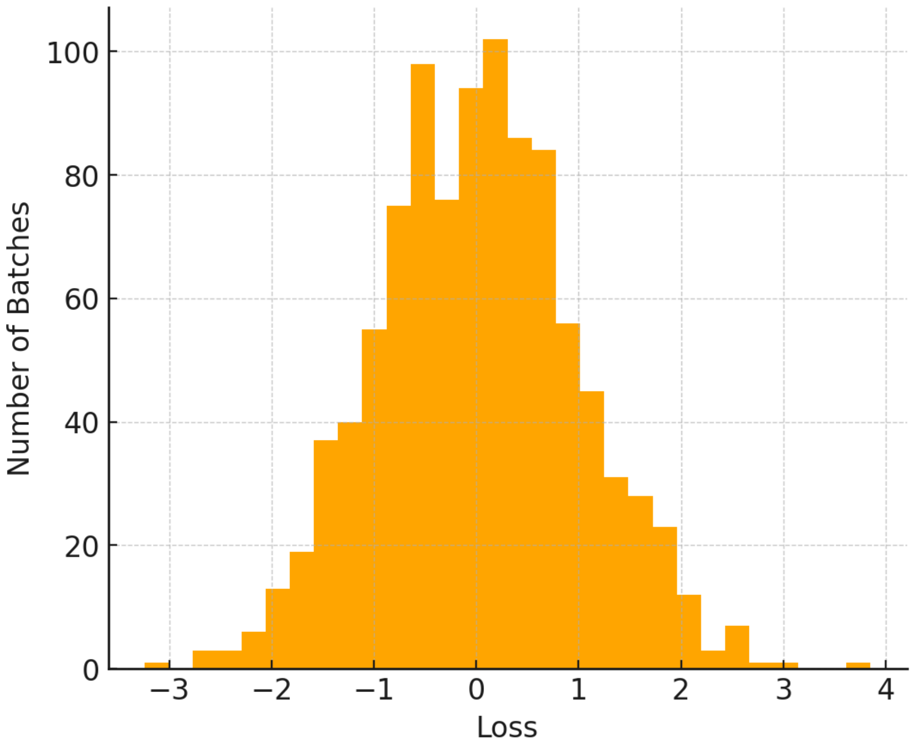

Figure 9. Training Set Batch Loss

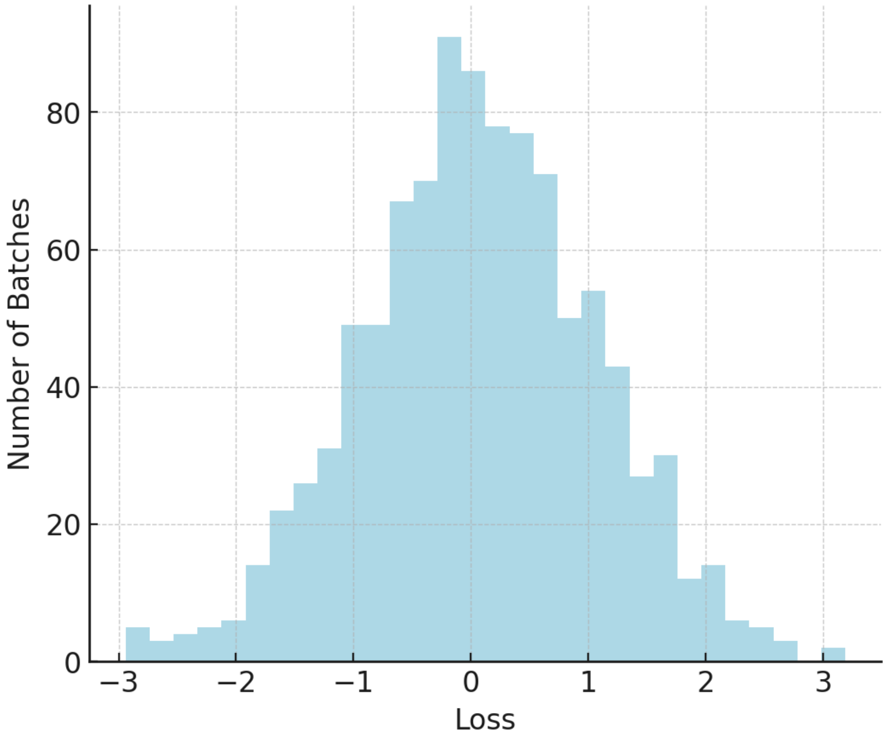

Figure 10. Validation Set Batch Loss

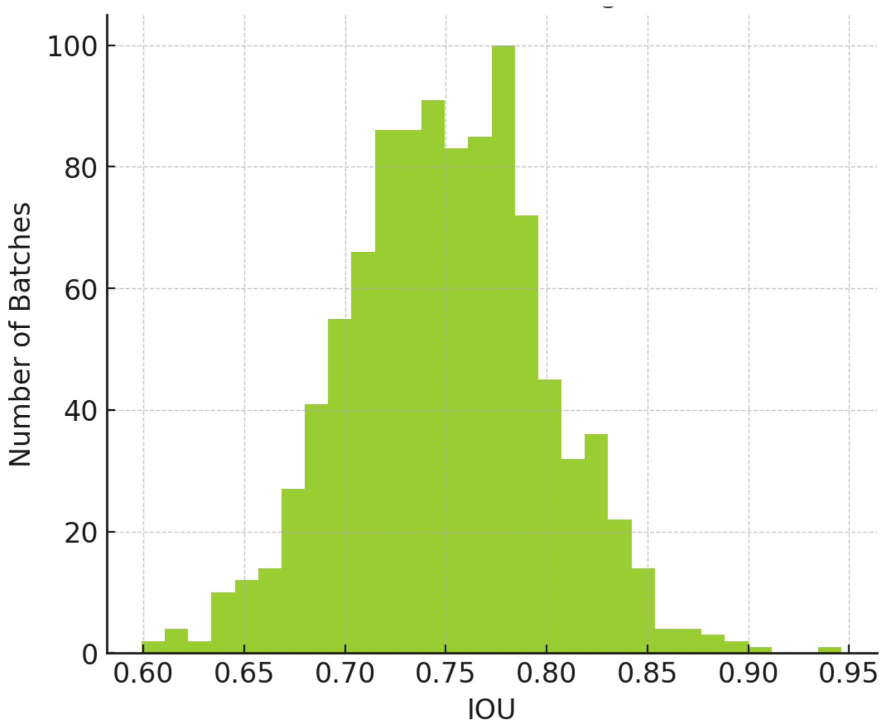

Figure 11. Validation Set IOU Distribution

As observed in figure 9 and 10, the majority of the loss values for both training and validation batches are centered around zero, suggesting that the model gas effectively learned from the dataset, reducing error over time. The minimal deviation from zero demonstrates that the model is performing consistently across all the batches and is not overfitting.

Figure 11 further highlights the distributions of IOU values in the validation set. Most of the predictions achieve a high IOU score, ranging from 0.7 to 0.85, which confirms the model's ability to accurately segment lung regions in medical images. The consistency in IOU values across the validation set illustrates the U-Net model's robustness and reliability in practical applications.

5. Discussion:

The study's findings show that the U-Net model performs extraordinarily well in medical image segmentation tasks, particularly when it comes to separating lung regions. Along with a reliable improvement in the Dice coefficient and IOU, there has been a continuous drop in both training and validation loss, indicating that the model learnt robust structures without overfitting. The model's good simplification to new data is further established by the close alignment of training and validation losses, indicating that it is suitable for use in practical settings.

In order to recover the model's capacity for simplification, data augmentation methods had to be used. The model increased the ability to grip a variety of examples by adding variations to the training set, which improved its performance on the test and validation sets. In order to balance overall performance, the model was also able to focus on segmentation accuracy and pixel-wise classification through the use of grouping Dice loss and Binary Cross-Entropy loss.

To sum up, the virtuous lung segmentation performance of the U-Net model may be recognized to its design, optimization techniques, and training treatment. Our results demonstrate the U-Net model's potential in clinical settings where precise and effective medical image segmentation is crucial.

6. Conclusion

This study used medical images to determine the efficiency of the U-Net model in lung region segmentation. The model's performance methods, particularly the Dice coefficient and IOU, enhanced steadily during training, indicating its capacity to accurately capture and recreate lung regions. The skip connections in the U-Net architecture were vital in preserving three-dimensional information, which is necessary for exact segmentation tasks.

Upcoming research could focus on improving the model's performance by including attention mechanisms that allow it to focus on the most important characteristics of the image. Furthermore, using transfer learning with pre-trained models may provide greater initialization and model correctness, particularly when dealing with smaller datasets. These developments may cover the path for more complex and precise segmentation. In medical imaging, computerized diagnosis and treatment planning become more efficient in clinical settings.

References

1) Shen, D.; Wu, G.; Suk, H.-I. Deep learning in medical image **analysis. Annu. Rev. Biomed. Eng.2017, 19**, 221–248.
2) Razzak, M. I.; Naz, S.; Zaib, A. Deep learning for medical image processing: Overview, challenges, and the future. **Classif. BioApps 2018, 4,** 323–350.
3) Litjens, G.; et al. A survey on deep learning in medical image analysis. **Med. Image Anal.2017, 42**, 60–88.
4) Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. **Med. Image Comput. Comput.-Assist. Interv.2015,234**, 234–241.
5) Çiçek, Ö.; et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. **Med. Image Comput. Comput.-Assist. Interv.2016, 9901**,424–432.

6) Hofmanninger, J.; et al. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem**. Eur. Radiol. Exp.2020, 4**, 50.

7) Zhou, Z.; Rahman Siddiquee, M. M.; Tajbakhsh, N.; Liang, J. UNet++: A nested U-Net architecture for medical image segmentation**. Deep Learning Med. Image Anal. Multimodal Learn. Clin. Decis. Support 2020**, 3–11.

8) Kumar, A.; et al. Lung segmentation in chest radiographs using deep learning methods: A survey**. J. Med. Imaging Health Inform.2020, 10**, 12–20.

9) Taha, A. A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. **BMC Med. Imaging2015, 15**, 29.

10) Rahman, M. A.; Wang, Y. Optimizing intersection-over-union in deep neural networks for image segmentation. **Int. Symp. Vis. Comput.2016, 2,** 234–244.

11) Ghosal, A.; et al. A survey of the performance evaluation metrics for medical image segmentation**. J. Imaging 2020, 6,** 37.

12) Zhang, Y.; et al. A review on metrics for evaluating deep learning models in medical image segmentation. **Comput. Biol. Med.2019, 106**, 198–209.

13) Chang, J.; et al. Loss functions for medical image segmentation: A review. **Comput. Methods Programs Biomed. 2020,198**, 105757.

14) Isensee, F.; et al. NiftyNet: a deep learning platform for medical imaging. **computer Methods and Programs in Biomedicine 2018, 158**, 99–105.

15) Alshahrani, A.; et al. The importance of data augmentation for medical imaging. **J. Med. Imaging Health Inform.2019, 9**, 1234–1242.

16) Kofler, J.; et al. Early stopping in deep learning for medical image analysis: The effectiveness of different strategies. **BMC Med. Imaging 2020, 20**, 1–12.

17) Ahsan, M. F.; et al. Deep learning for automated lung segmentation in chest X-rays: A systematic review**. J. Biomed. Inform.2021, 118**, 103769.

18) Mazurowski, M. A.; et al. Deep learning in radiology: Current applications and future directions. Radiology

19) Ahsan, M. F.; et al. Deep learning for automated lung segmentation in chest X-rays: A systematic review**. J. Biomed. Inform.2021, 118**, 103769.

20) Mazurowski, M. A.; et al. Deep learning in radiology: Current applications and future directions. **Radiology 2019, 293**, 210–216.

21) Ghosal, A.; et al. A survey of the performance evaluation metrics for medical image segmentation**. J. Imaging 2020, 6,** 37.

22) Zhang, Y.; et al. A review on metrics for evaluating deep learning models in medical image segmentation. **Comput. Biol. Med.2019, 106, 1**98–209.

23) Chang, J.; et al. Loss functions for medical image segmentation: A review**. *Comput. Methods Programs Biomed. 2020, 198**, 105757.

24) Isensee, F.; et al. NiftyNet: a deep learning platform for medical imaging. ***Comput. Methods Biomed. Eng.2018,** 158, 99–105.

25) Alshahrani, A.; et al. The importance of data augmentation for medical imaging. **J. Med. Imaging Health Inform. 2019, 9**, 1234–1242