International Journal of Innovative Research and Scientific Studies, 8(1) 2025, pages: 380-393

IJIRSS
International Journal of Innovative Research & Scientific Studies

**ISSN:** 2617-6548

URL: www.ijirss.com

# Neural network solutions for artificial intelligence based on the new MIU-Net model for segmentation of the lung images in the diagnosis and treatment of lung diseases

Gulnaz Nabiyeva[1*], Aliya Kalizhanova[2], Nurgul Abutalipova[3], Dina Koishiyeva[4], Sultan Akhmetov[5]

[1]*S. Asfendiyarov Kazakh National Medical University, Almaty, Kazakhstan, and Heriot-Watt University, Edinburgh, UK.*
[2]*Institute of Information and Computer Technologies CS MSHE RK, Almaty University of Energy and Communications named after G.Daukeev, Almaty, Kazakhstan.*
[3]*Abai Kazakh National Pedagogical University, Almaty, Kazakhsta, and Heriot-Watt University, Edinburgh, UK.*
[4]*Almaty University of Energy and Communications named after G.Daukeev, Almaty, Kazakhstan.*
[5]*Al-Farabi Kazakh National University, Almaty, Kazakhstan.*

Corresponding author: Gulnaz Nabiyeva (*Email: gulnaz_nc@mail.ru*)

## Abstract

The purpose of this research is to segment the chest on X-rays, a crucial step in medical image processing pivotal for diagnosing and treating lung diseases. By employing modifications to the U-Net model, this study endeavors to enhance chest segmentation on X-rays. The approach includes introducing alterations to the basic architecture of the U-Net model, integrating Inception blocks, and a using a squeeze-and-excitation mechanism to improve segmentation accuracy. The Shenzhen dataset, comprising chest radiographs, serves as the subject of investigation, highlighting the practical application of these modifications. The utilization of the MIU-Net model for automatic chest organ segmentation underscores its significance in the realm of lung disease diagnosis and treatment. Experimental methodologies encompass two types of data augmentation: Contrast Limited Adaptive Histogram Equalization (CLAHE) and the introduction of Gaussian noise, to test the model's robustness under various conditions. A comparative analysis is conducted against both the baseline U-Net and U-Net with Inception blocks. The results show that the improved U-Net model that includes Inception and Squeeze-and-Excitation (SE) is a lot better than the original U-Net. Specifically, the Dice coefficient for the improved model stands at 0.9040 for the original data, 0.9306 with CLAHE application, and 0.9232 with Gaussian noise addition. These findings underscore the importance of the research, emphasizing the significance of improving the accuracy of chest segmentation on X-rays for early disease detection and treatment optimization, which are the practical implications of this study. The research's implications extend beyond academic interest, offering potential enhancements in clinical practices for lung disease management.

**Keywords:** Augmentation, Inception block, Medical image processing, Segmentation, Squeeze and excitation, U-Net model, X-Ray.

**DOI:** 10.53894/ijirss.v8i1.4171
**Funding:** This research is supported by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant number: AP 19679153).
**History: Received:** 6 February 2024/**Revised:** 11 December 2024/**Accepted:** 27 December 2024/**Published:** 17 January 2025

**Competing Interests:** The authors declare that they have no competing interests.
**Authors' Contributions:** All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
**Transparency:** The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
**Institutional Review Board Statement:** The Ethical Committee of the Institute of Information and Computational Technologies CS MSHE RK, Kazakhstan has granted approval for this study on 3 August 2023 (Ref. No. 298/23-25).
**Publisher:** Innovative Research Publishing

## 1. Introduction

Lung diseases, which include diseases such as lung cancer, tuberculosis, and pneumonia, represent a major global health challenge. These respiratory diseases are not only widespread, but they have a significant impact on people's well-being and lives. Lung cancer stands as the foremost contributor to global cancer incidence and mortality, responsible for a staggering estimated 2 million diagnoses and 1.8 million deaths annually [1]. The intricate interplay of risk factors, spanning tobacco use, environmental exposures, and genetic predispositions, underscores the intricate nature of lung cancer. Chest radiography, due to its affordability, actively diagnoses lung diseases such as pneumonia, tuberculosis, cancer, and more. Manual analysis of images by expert radiologists can take more time, and artificial intelligence applications can help experts in diagnosis [2]. In addition, advances in imaging technologies and machine learning algorithms continue to change the landscape of respiratory health diagnostics, offering new opportunities for early detection and personalized treatment strategies [3]. Computer-aided diagnosis (CAD) systems based on convolutional neural networks can help in medical recognition tasks as well as in highlighting the region of interest in images [4]. These systems, effective in segmenting complex lung structures, utilize CNN architectures such as U-Net [5] due to their exceptional feature extraction capabilities. U-Net, known for its symmetric expanding and contracting contours, has become the standard in medical image segmentation [6]. Lung segmentation on X-ray images, is a challenging task, with difficulties caused by overlapping anatomical structures, variability in image quality depending on imaging conditions and patient, and the presence of noise and artefacts [7]. Researchers and the developers are actively working to improve segmentation algorithms due to the complexity and increasing volume of data [8]. Residual U-Net is a modification of the traditional U-Net architecture in which residual links or residual connections are integrated [9]. In Singadkar, et al. [10] a deeper model of this architecture was developed for lung segmentation. The Dense Inception U-Net model with convolutional layers, employs an additional Inception module and dense layers [11]. Also, given the limited amount of annotated data, it is necessary to use a variety of augmentation techniques to expand the dataset. In studies [12], various data augmentations have been applied, aiming not only to extend the datasets but also to bring the training samples closer to real clinical conditions [13]. In our study, augmentation techniques were applied to the X-ray images, such as CLAHE [14] for contrast enhancement and Gaussian blurring [15].

Our study proposes an improved version of the U-Net model, incorporating additional filters for deeper image processing. The integration of Inception blocks provides multi-scale analyses [16] while the addition of squeeze and excitation modules facilitates the extraction of key features [17].

## 2. Method and Materials

The proposed MI-UNet model, Figure 3, based on the U-Net architecture, includes additional blocks for semantic segmentation. The encoder performs gradual image compression to extract key features, and the decoder reconstructs image dimensions for detailed segmentation. The inclusion of Inception blocks provides multi-level feature analysis, and SEblocks drill down into the features.

### *2.1. U-Net Architecture*

The U-Net architecture is the basis for many state-of-the-art models in the task of biomedical image segmentation. It is based on convolutional layers. The MIU-Net model uses 3x3 convolutional blocks with kernels in the encoder, where the number of filters starts at 32 and doubles at each subsequent layer. The decoder uses 2 x 2 transposed convolutions [18] to increase the size of the feature maps, with the number of filters decreasing in reverse order. Moreover, the feature maps created by the transposed convolutional layers are high-resolution representations of the input data. This feature allows the network to make accurate localization and classification decisions, which are essential in biomedical image analysis.

### *2.2. Squeeze Excite Block*

The squeeze-excite block Figure 1, function in the proposed model implements the squeeze-and-excite mechanism shown in Figure 1, which is a form of feature reconfiguration used to improve convolutional neural networks. This unit works by first applying global average integration to the input tensor, reducing each channel to a single scalar value. This

compression operation captures the global spatial information of the input signal [19]. The SE block applies a series of convolutional filters to the input tensor, systematically recalibrating the feature maps. This adaptive recalibration ensures that the model pays more attention to the most informative features. This operation would be represented as Formula 1.

$$u_c = v_c * X = \sum_{s=1}^{C} v_c^s * x^s \quad (1)$$

Where * represents the convolution operation. We convolve each filter with the corresponding channel of the input tensor X, than aggregate the results to form the final output feature map. This technique is designed to enhance the neural network's sensitivity and its ability to discern pertinent patterns within the data, thereby bolstering the overall predictive performance of the model [20].

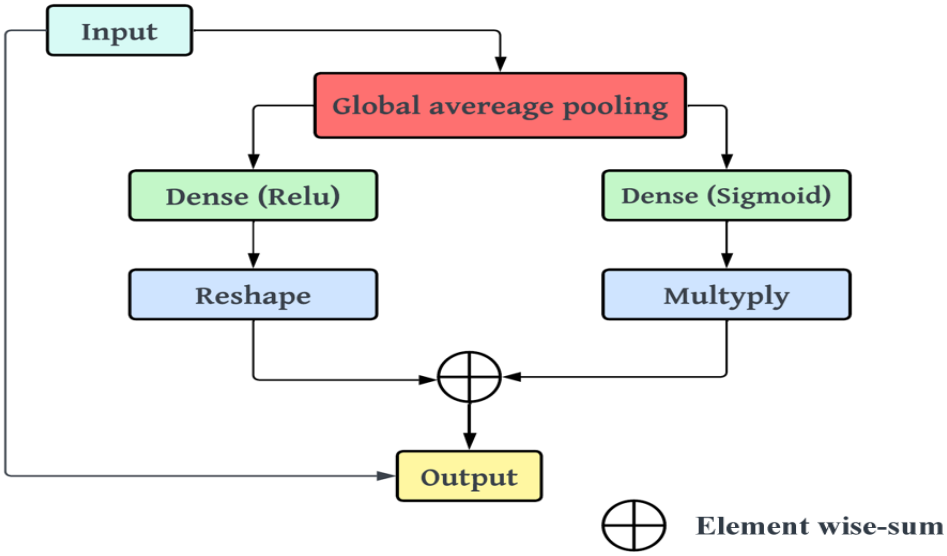


**Figure 1.**
Squeeze excite block.

The resulting vector then passes through two fully connected dense layers, the first of which uses the ReLU activation function for nonlinearity to reduce the channel sizes by a given ratio, and the second of which use a sigmoidal activation function to generate weighs on the channels to restores the sizes to the original number of channels [21].

*2.3. Inception Block*

The inception block Figure 2, Function in the proposed model, builds a composite convolution block in the neural network by applying parallel 1x1, 3x3, and 5x5 convolution paths to the input tensor, each capturing features at different scales, and a maximal merge operation followed by 1x1 convolution to reduce dimensionality.

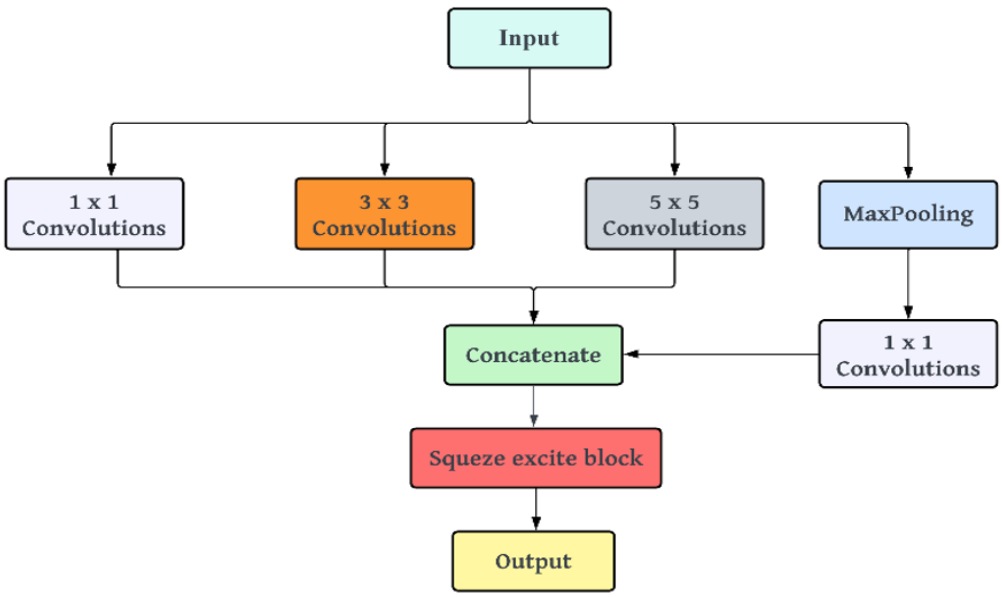


**Figure 2.**
Inception block.

These paths are then combined to merge their feature maps, and the merged output is fed into a squeeze-excite block that applies channel-by-channel feature recalibration to emphasize informative features and suppress less useful ones. We employ the maximum merge operation to integrate the distinct feature maps into a unified tensor, following the parallel processing paths.

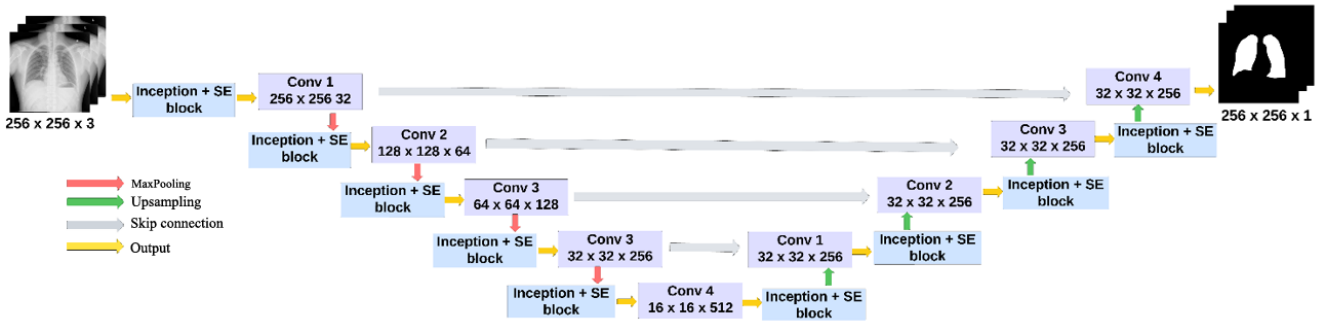


**Figure 3.**
MIU-Net model architecture.

The inclusion of 1x1 convolutions, in addition to their role in feature extraction, has implications for dimensionality reduction. They act as narrow layers that compress the feature space, thereby reducing the computational and parametric growth that would arise from directly merging multi-scale feature maps. Figure 3 represents the proposed architecture of MIU-Net with embedded SE and Inception blocks.

*2.4. Improving Generalization Ability*

Because the model architecture has a complex layer structure with additive blocks, the L2 regularizer [22], in convolutional layers, was additionally introduced to prevent overtraining. Without regularization, the model attempts to minimize the loss function without considering the complexity of the model, which can lead to high sensitivity to training data. The regularizer adds a loss function penalty proportional to the magnitude of the weights of these layers. The regularized loss function comprises the space between the forecasted and actual target labels and the L2 of the vector of weights according to Formula 2.

$$L(w) = \|\hat{y} - y\|_2^2 + l\,\|w\|_2^2 \qquad (2)$$

The regularization factor is set at 0.0001, which is equal to 1e-4, this factor can be adjusted depending on the complexity of the problem and the size of the training dataset [23].

*2.5. Comparison of MIU-Net with Experimental models*

To compare the proposed MIU-Net model, the algorithms of the basic U-Net model were used in the experiment, as well as the IU-Net model with the inclusion of an additional block. The basic U-Net architecture uses dual convolutional layers with 128 and 256 filters in the encoder and 512 in the narrow layer, augmented with packet normalization layers to improve training stability. The model also incorporates max-pooling for dimensionality reduction and up-sampling to restore image dimensions in the decoder. In contrast, IU-Net extends the U-Net architecture by integrating Inception blocks into each encoder and decoder layer, which allows for deeper learning of images by utilizing parallel convolutional paths with different kernel sizes 1x1, 3x3, 5x5, and max-pooling. Table 1 shows the basic blocks and a description of the operation and parameters of the MIU-Net model filters.

**Table 1.**
Architecture of the MI-U-Net.

| Block | Operation | Layers parameters |
|---|---|---|
| Encoder block 1 | Inception block + SE block + Convolution + Activation + BatchNormalization + Dropout + MaxPooling | Inception: Filters: 32, SE: Compression ratio 8, dropout: 0.2, L2 regularization: 1e-4 |
| Encoder block 2 | Inception block + SE block + Convolution + Activation + BatchNormalization + Dropout + MaxPooling | Inception: Filters: 64, SE: Compression ratio 8, dropout: 0.2, L2 regularization: 1e-4 |
| Encoder block 3 | Inception block + SE block + Convolution + Activation + BatchNormalization + Dropout + MaxPooling | Inception: Filters: 128, SE: Compression ratio 8, dropout: 0.2, L2 regularization: 1e-4 |
| Encoder block 4 | Inception block + SE block + Convolution + Activation + BatchNormalization + Dropout + MaxPooling | Inception: Filters: 256, SE: Compression ratio 8, dropout: 0.2, L2 regularization: 1e-4 |

| Block | Operation | Layers parameters |
|---|---|---|
| Bridge | Convolution + Activation + BatchNormalization + Dropout | Filters: 512, dropout: 0.2, L2 regularization: 1e-4 |
| Decoder block 1 | Transposed convolution + Concatenation + Inception block + SE block + Convolution + Activation + BatchNormalization + Dropout | Transposed convolution: Filters: 256, Inception: Filters: 256, SE: Compression ratio 8, L2 regularization: 1e-4 |
| Decoder block 2 | Transposed convolution + Concatenation + Inception Block + SE block + Convolution + Activation + BatchNormalization + Dropout | Transposed convolution: Filters: 128, Inception: Filters: 128, SE: Compression ratio 8, L2 regularization: 1e-4 |
| Decoder block 3 | Transposed convolution + Concatenation + Inception block + SE block + Convolution + Activation + BatchNormalization + Dropout | Transposed convolution: Filters: 64, Inception: Filters: 64, SE: Compression ratio 8, L2 regularization: 1e-4 |
| Decoder block 4 | Transposed convolution + Concatenation + Inception Block + SE block + Convolution + Activation + BatchNormalization + Dropout | Transposed convolution: Filters: 32, Inception: Filters: 32, SE: Compression ratio 8, L2 regularization: 1e-4 |
| Output | Convolution + Activation | Filters: 1, Activation: Sigmoid |

Table 1 summarizes the architecture of the MIU-Net convolutional neural network, detailing the configuration of the encoder and decoder blocks, bridge, and output layer for image segmentation tasks. Each encoder block uses an initial block for multiscale feature extraction, a compression and excitation SEblock for adaptive feature recalibration, followed by convolution, activation, batch normalization, dropout for regularization, and maximum pooling for spatial dimensionality reduction, with filter sizes doubling from 32 to 256 throughout all four blocks. The bridge then processes the features using convolution and dropout.

Decoder blocks repeat the structure of the encoder, but use transposed convolutions to upsample while maintaining the same filter sizes and SE compression ratios, and include concatenation steps to combine the features received from the encoder.

Table 2 shows a comparison of the size and number of parameters for the U-Net, IU-Net, and MIU-Net models.

**Table 2.**
Model size and parameters.

| Model | Size (MB) | Parameters (M) |
|---|---|---|
| U-Net | 28.73 | 7,531,521 |
| IU-Net | 9.91 | 2,598,209 |
| MIU-Net | 167.64 | 43,945,921 |

The model size, expressed in megabytes (MB), indicates the amount of memory occupied by each model, while the number of parameters, expressed in millions (M), reflects the depth of each network. In addition, the scalability of these models to different hardware platforms, from high-performance GPUs in research centers to more modest computing resources in clinical settings, is directly related to their size and complexity. According to Table 2, U-Net is the smallest and least complex model, while MIU-Net is significantly larger and contains more parameters, indicating its high complexity and advanced functionality. IU-Net occupies an intermediate position in terms of size and complexity.

*2.6. Shenzhen Dataset Description and Augmentation*

The study utilized a dataset [24] consisting of 566 x-ray images and 566 corresponding chest masks, each with a resolution of 512 x 512 pixels. The American national Library of Medicine and No. 3 Hospital in Shenzhen, China, collaborated to obtain these images.

This dataset is distinctive in that it includes both healthy chest X-rays and images with evidence of tuberculosis. Figure 4 shows an example of an image and mask from this dataset.

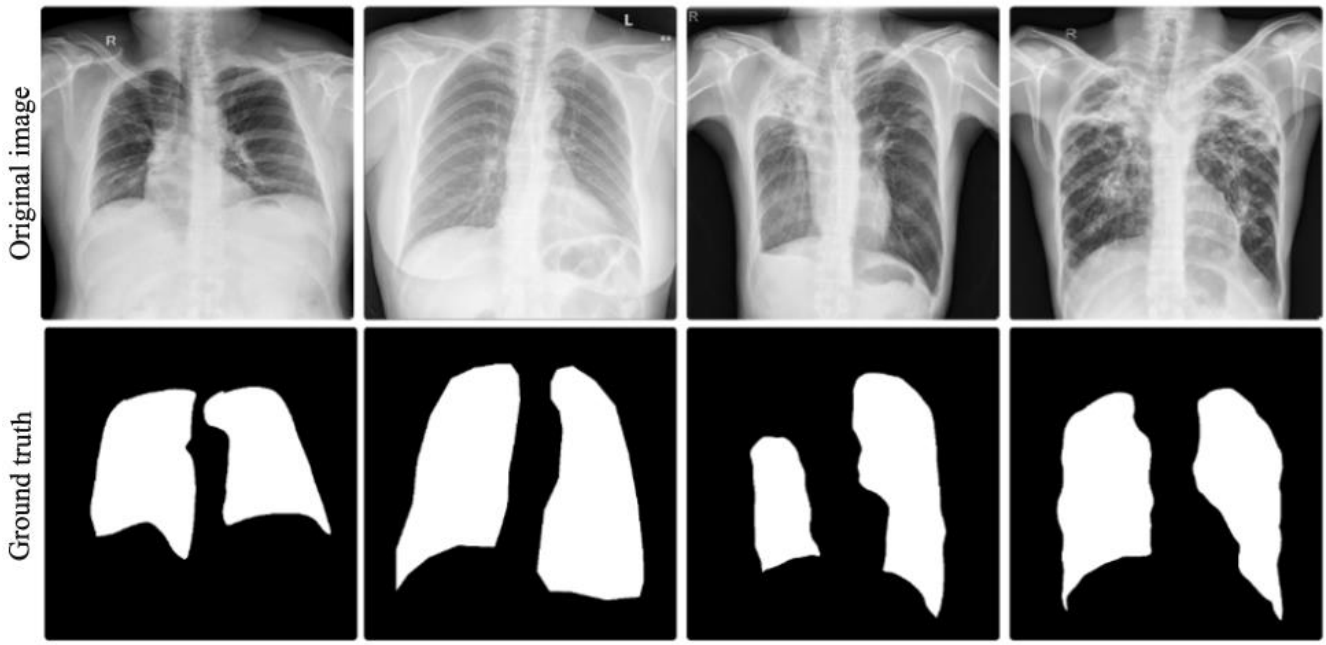


**Figure 4.**
Some samples from the original dataset.

*2.6.1. Augmentation Methods*

In the experiment, two augmentation strategies were implemented, in which the number of images and masks was increased from 566 to 800. The first strategy involved a combination of transformations, including randomly varying brightness and contrast, applying CLAHE to improve contrast, random 90-degree rotation, combined shift, zoom, and rotation with constraints, and rotation within a specified range of angles. The second strategy involved the application of Gaussian blur to add artificial noise, imitation effects due to patient movement, or equipment deficiencies. The experiment uses these methods to not only make the model work better, but also to make sure that these improvements have real effects in clinical practice, especially when it comes to correctly separating lung x-rays in a variety of imaging situations. Figure 5 shows some example data after augmentation with CLAHE, and Figure 6 shows sample images with Gaussian blurring.

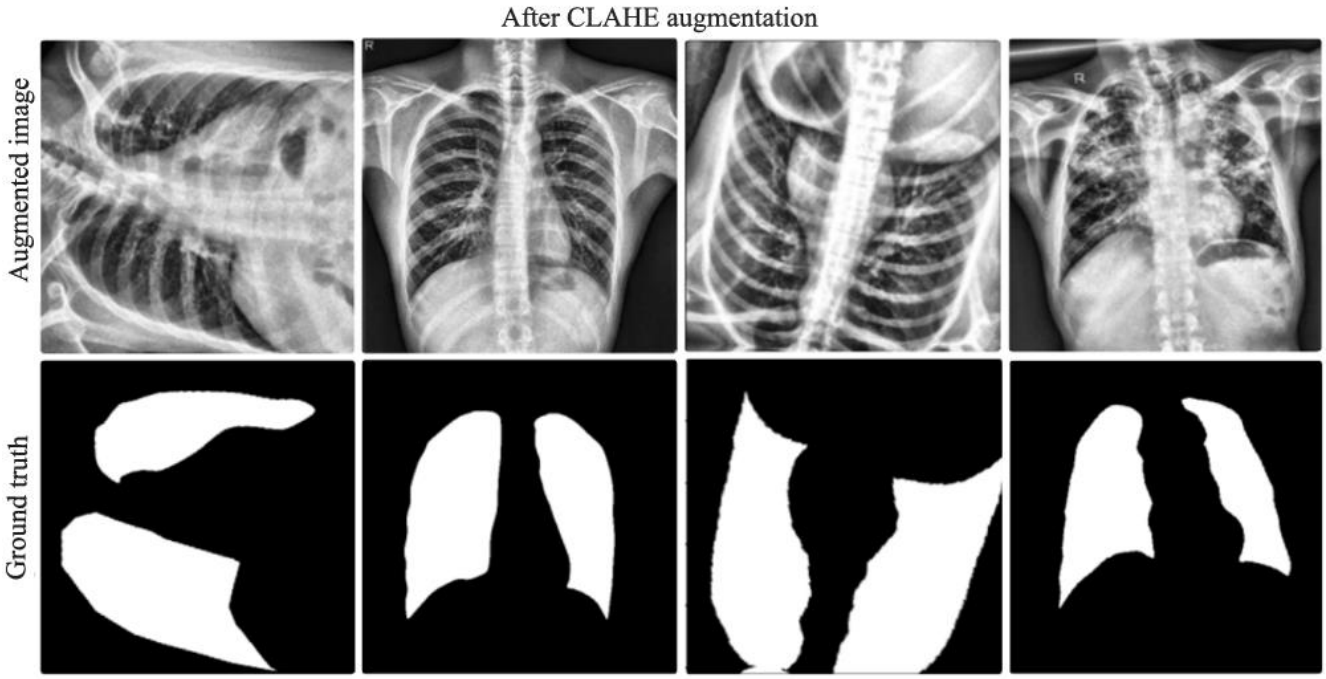


**Figure 5.**
Several examples after augmentation with CLAHE.

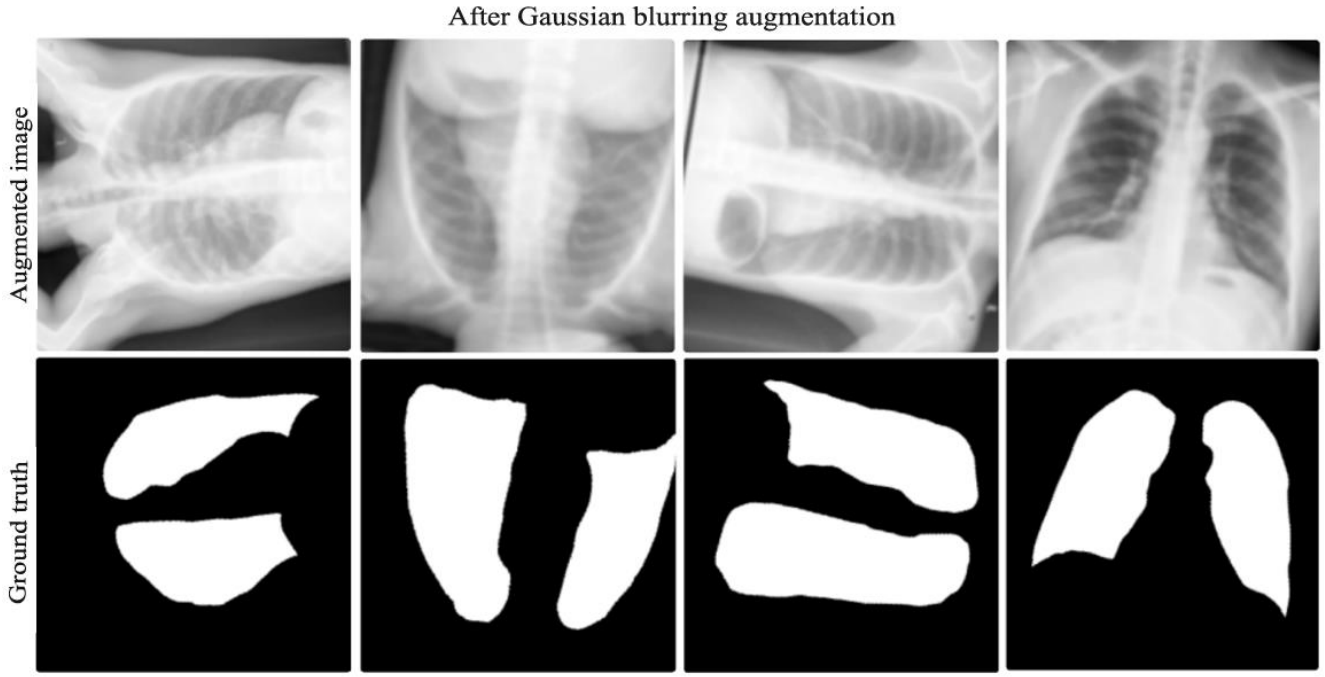


**Figure 6.**
Several examples after augmentation with Gaussian blurring.

The use of CLAHE and Gaussian blur not only increases the variability of the dataset but also simulates real-world conditions that models may be faced with during processing.

## 3. Implementation Detail

The study was conducted in Google Colab environment [25] using Python programming language version 3.10.12 and the TensorFlow library version 2.15.0. A Tesla T4 GPU with driver version 535.104.05 and CUDA version 12.2 and 50.99 GB of RAM was used for data processing. Adam, with a learning rate of 1e-4 and a patch size of 2 was used as an optimizer [26].

## 4. Evaluation Metrics

The proposed MIU-Net model uses several key evaluation metrics to estimate and compare its performance with both the baseline, U-Net and the U-Net model for lung segment allocation. In this section, the key metrics are described and explained, including binary cross entropy loss (BCE loss),dice coefficient denoted as dice, and mean intersection over union, (Mean IOU). Below we present their equations and definitions [27, 28].

$$BCE\ loss(x,y) = \sum -(y * log(y) + (1-y)\, log(1-y)) \quad (3)$$

BCE loss is a key measure in medical image segmentation, quantifying the accuracy of pixel-level binary classifications. It penalizes the model for any deviation between the predicted probabilities and the actual binary labels in the images [25].

$$Dice(A,B) = 2\frac{|A| \cap |B|}{|A|+|B|} \quad (4)$$

This metric defines the similarity between two sets, A and B, by measuring twice the intersection of the sets divided by the sum of their sizes [26].

$$Mean\ IoU = \frac{|mask \cap prediction|}{|mask \cup prediction|} \quad (5)$$

Mean IOU computes the overlap between the predicted segmentation and the real mask divided by their union, effectively measuring prediction accuracy in segmentation tasks [29].

## 5. Experiment and Results of Study

This section presents the experimental results comparing the proposed MIU-Net model with the baseline U-Net and IU-Net to valid data from the original Shenzhen dataset. We extend the evaluation to the datasets using augmentation techniques, introducing CLAHE in the second subsection and Gaussian blurring in the third.

### *5.1. Performance Evaluation on the Original Dataset*

The initial dataset consists of 566 image samples, each accompanied by its corresponding annotations. Before training, the images were resized from 512 to 256 pixels and then segmented into training, validation, and test sets containing 452, 56, and annotated masks, respectively. This pre-processing was aimed at optimizing computational efficiency and ensuring a balanced distribution of data across subsets. The output values of the segmentation results of the underlying metrics are shown in Table 3. We have rounded the values to four decimal places.

**Table 3.**
Segmentation results on the original dataset.

| Model | BCE loss | Mean IOU | Dice |
|---|---|---|---|
| U-Net | 0.6079 | 0.4493 | 0.6185 |
| IU-Net | 0.0837 | 0.9035 | 0.9379 |
| MIU-Net | 0.1003 | 0.9040 | 0.9495 |

The results show that U-Net achieves moderate performance, with BCE loss at 0.6079, Mean IOU at 0.4493, and Dice at 0.6185. In contrast, IU-Net and MIU-Net show higher accuracy, with IU-Net showing BCE loss at 0.0837, Mean IOUat 0.9035, and Dice at 0.9379, and MIU-Net showing BCE loss at 0.1003, Mean IOU at 0.9040, and Dice at 0.9495. These results indicate that IU-Net and MIU-Net outperform U-Net in terms of accuracy in identifying regions of interest on this particular dataset. The validation curves for the Dice and Loss metrics are shown in Figure 7 and Figure 8.

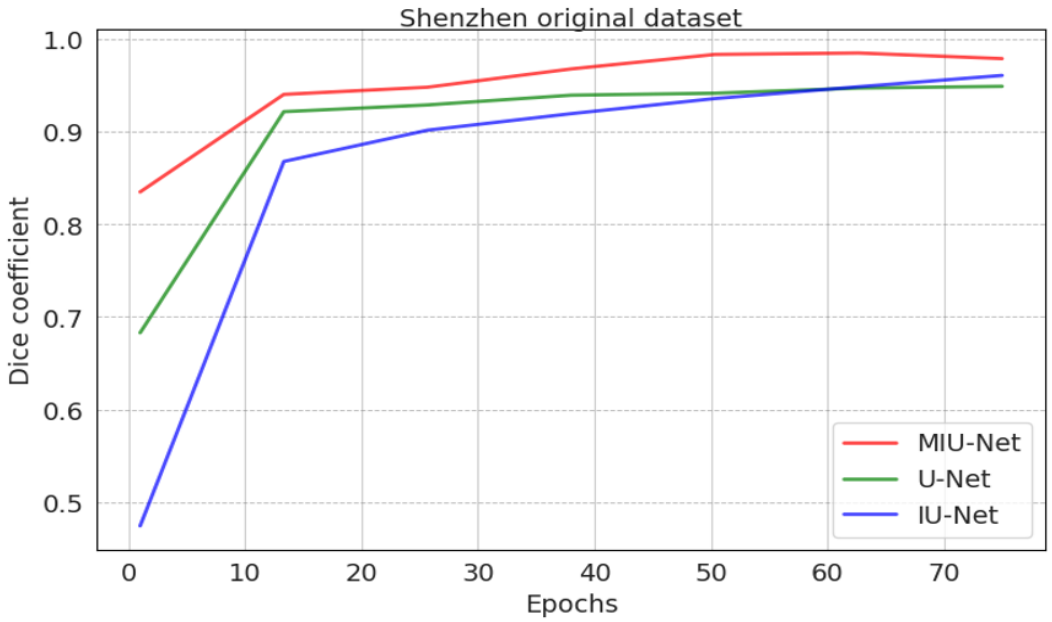


**Figure 7.**
Dynamics of dice coefficient on the original dataset.

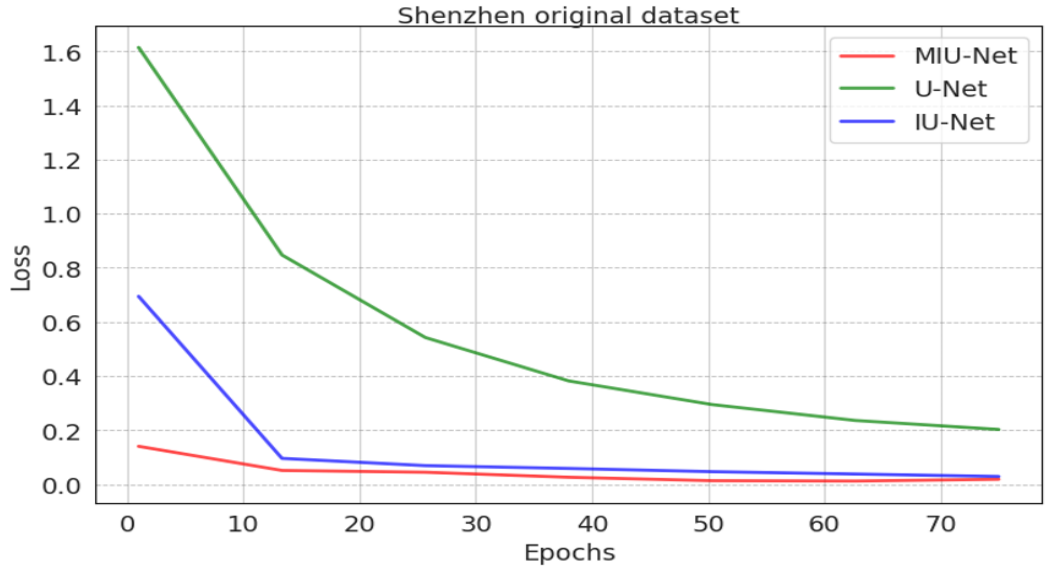


**Figure 8.**
Dynamics of BCE loss on the original dataset.

Figure 9 shows the segmentation results on the original dataset, where the differences in segmentation quality achieved by the different models can be clearly seen.

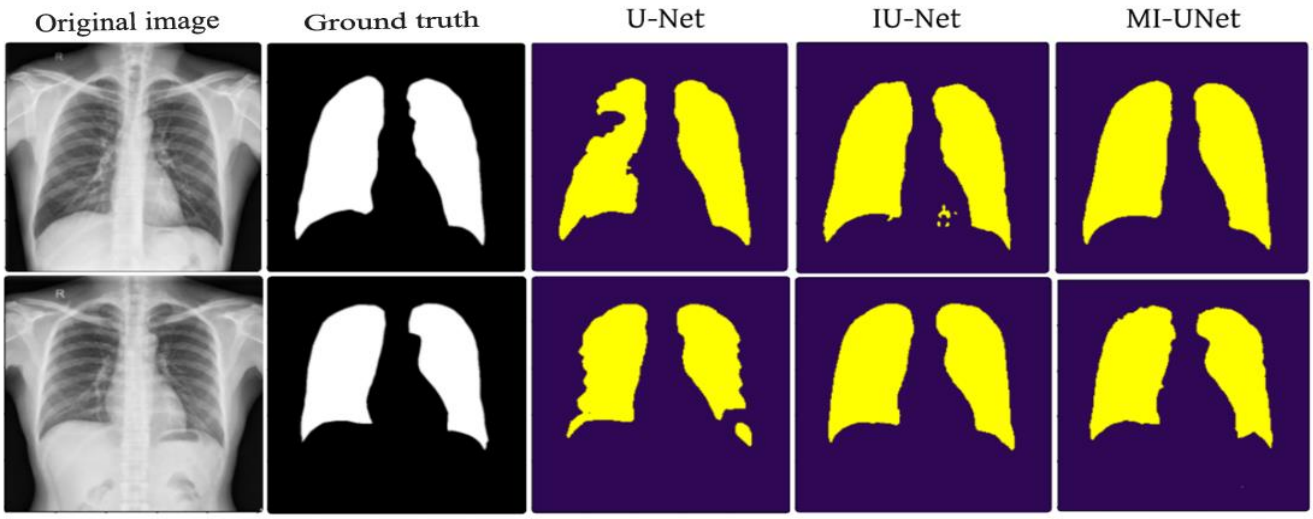


**Figure 9.**
Segmentation results on the original dataset.

*5.2. Performance Evaluation after CLAHE Augmentation*

The use of CLAHE implied not only an improvement in the visual clarity of the images, but also a potential improvement in the accuracy of the trained models. Table 4 provides the results of the model's performance on the validation data.

**Table 4.**
Segmentation results with CLAHE augmentation.

| Model | BCE loss | Mean IOU | Dice |
|---|---|---|---|
| U-Net | 0.1436 | 0.8847 | 0.9286 |
| IU-Net | 0.1184 | 0.8888 | 0.9354 |
| MIU-Net | 0.1880 | 0.8953 | 0.9449 |

Applying the CLAHE method resulted in a significant improvement in the performance of all three models. The U-Net model showed improvement with BCE loss of 0.1436, Mean IOU of 0.8847 and Dice of 0.9286, indicating an improvement in overall segmentation accuracy. The IU-Net and MIU-Net models performed better, with IU-Net achieving a BCE loss of 0.1184, Mean IOU of 0.8888 and Dice of 0.9354, and MIU-Net achieving a BCE loss of 0.1880, Mean IOU of 0.8953 and Dice of 0.9449. Below in Figure 10 and 11 is a diagram showing the comparative trend, and Figure 12 shows the lung segment predictions.

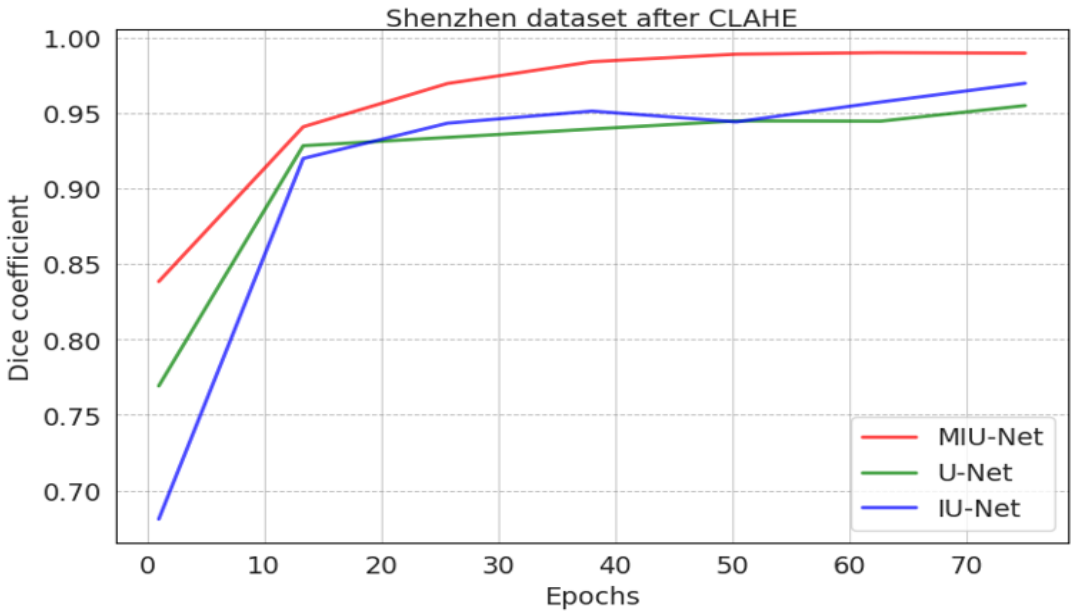


**Figure 10.**
Dynamics of Dice coefficient after CLAHE augmentation.

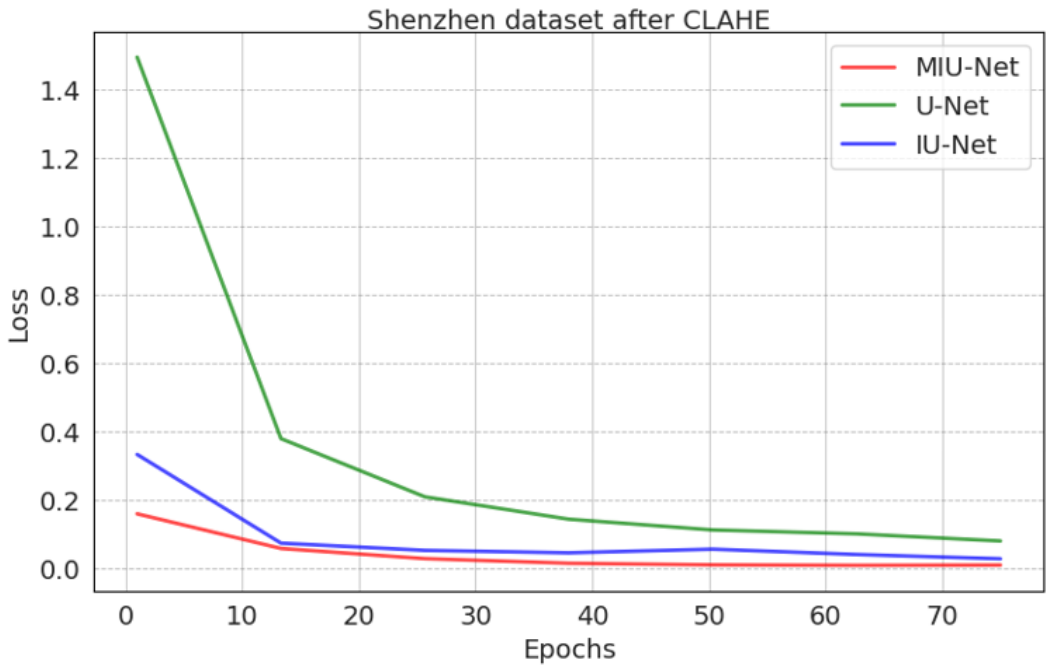


**Figure 11.**
Dynamics of BCE loss after CLAHE augmentation.

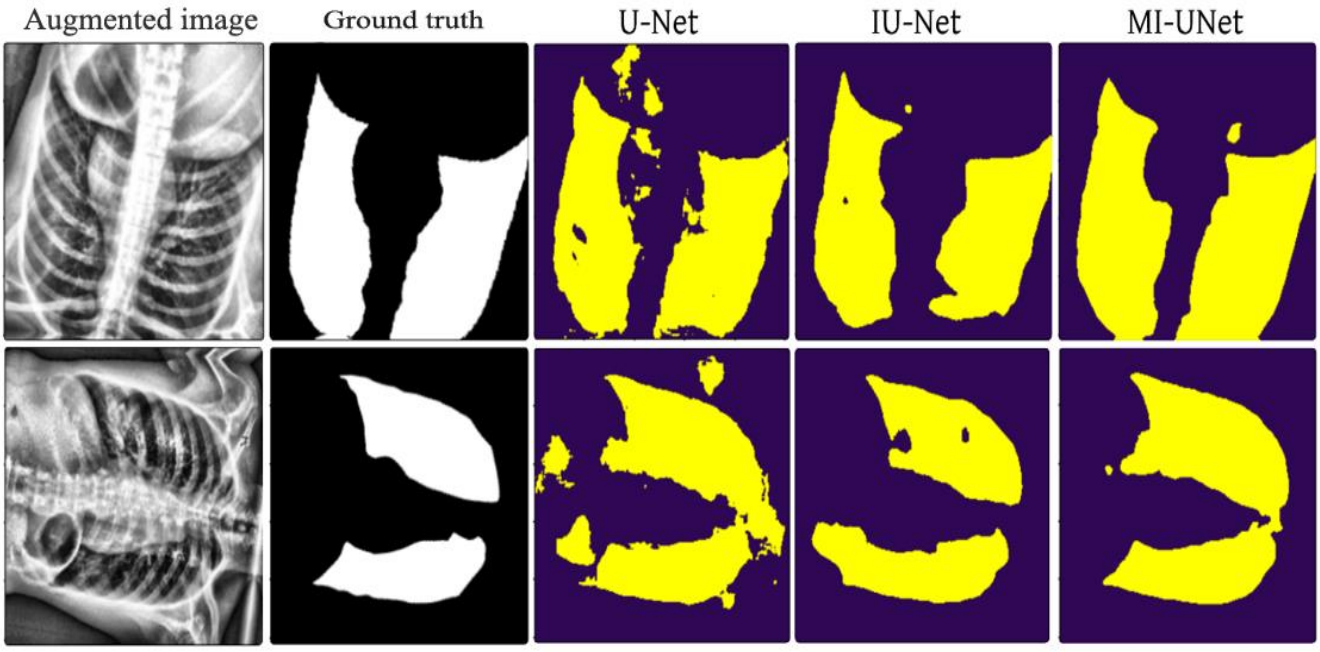


**Figure 12.**
Segmentation results after augmentation with CLAHE.

Figure 12 shows the forecasting models results' and despite improved results on key indicators, there are still some errors in the output of segmentation forecasting. This highlights that even with improved accuracy of indicators such as BCE loss, Mean IOU, and Dice coefficient, models can still allow inaccuracies in identifying and selecting target segments. Various factors, such as limitation in the model architecture, data complexity, or flaws in the learning process, can cause such errors.

### *5.3. Performance Evaluation After Gaussian Blurring*

The implementation of Gaussian blur as an augmentation method in our experiment was motivated by the need to train models for realistic imaging conditions typical of the medical field. Contrary to high-frequency noise, which can negatively affect image quality by masking important details, the low-frequency noise produced by Gaussian blur reproduces image blurring due to patient motion, equipment vibration, or other technical features. Further, Table 5 shows the numerical results of the learning outcomes.

**Table 5.**
Segmentation results with Gaussian blurring augmentation.

| Model | BCE loss | Mean IOU | Dice |
|---|---|---|---|
| U-Net | 0.1699 | 0.8945 | 0.8945 |
| IU-Net | 0.1131 | 0.9131 | 0.9361 |
| MIU-Net | 0.1470 | 0.8722 | 0.9131 |

According to segmentation results after augmentation with Gaussian blur, there are differences in the performance of the U-Net, IU-Net and MIU-Net models. The U-Net model shows average performance with a BCE loss of 0.1699 and Mean IOU and Dice coefficient of 0.8945, indicating its moderate adaptation to the changes made. In contrast, the IU-Net and MIU-Net models perform significantly better: IU-Net with BCE loss 0.1131, Mean IOU 0.9131 and Dice 0.9361, and MIU-Net with BCE loss 0.1470, Mean IOU 0.8722 and Dice 0.9131. These findings indicate that both of these models are more effective for segmentation tasks in low-frequency noise environments. Following Figure 13, Figure 14, and Figure 15 are plots and visual lung segment predictions illustrating the effect of Gaussian blurring on segmentation performance.

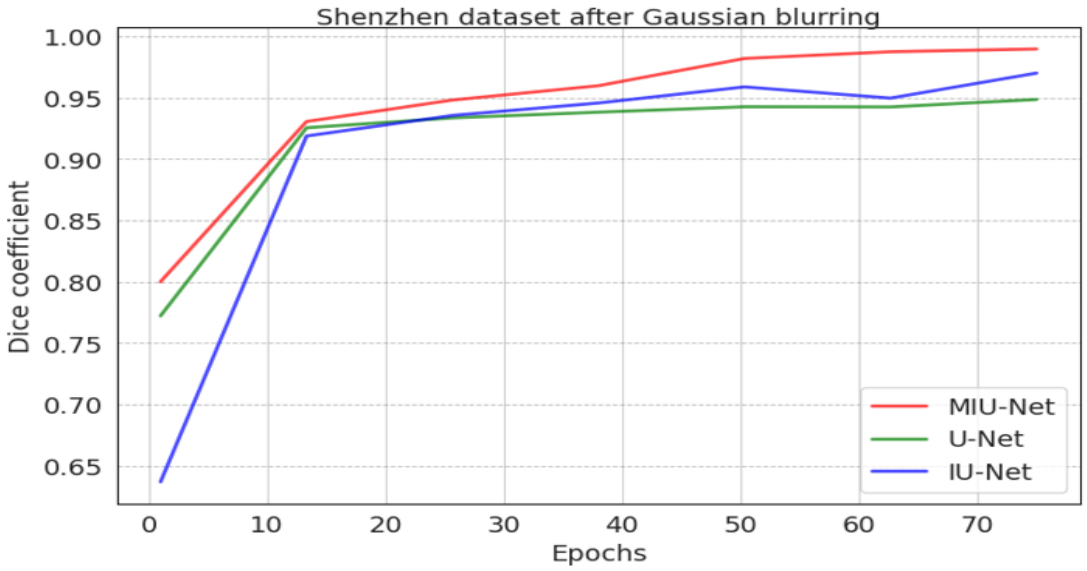


**Figure 13.**
Dynamics of dice coefficient after Gaussian blurring augmentation.

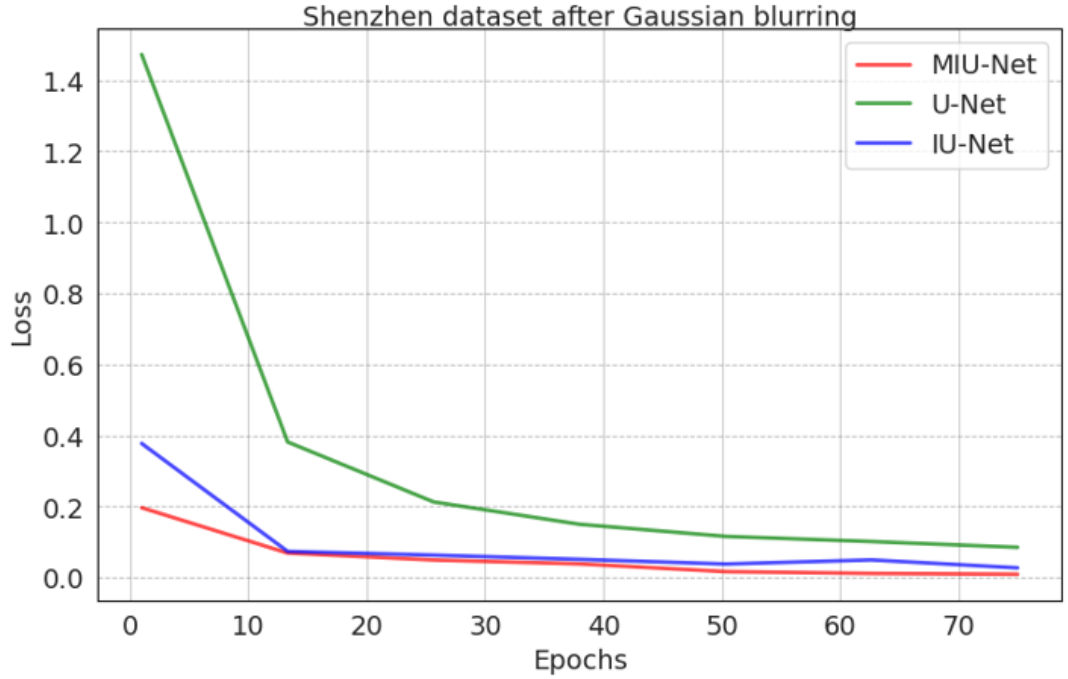


**Figure 14.**
Dynamics of BCE loss after Gaussian blurring augmentation.

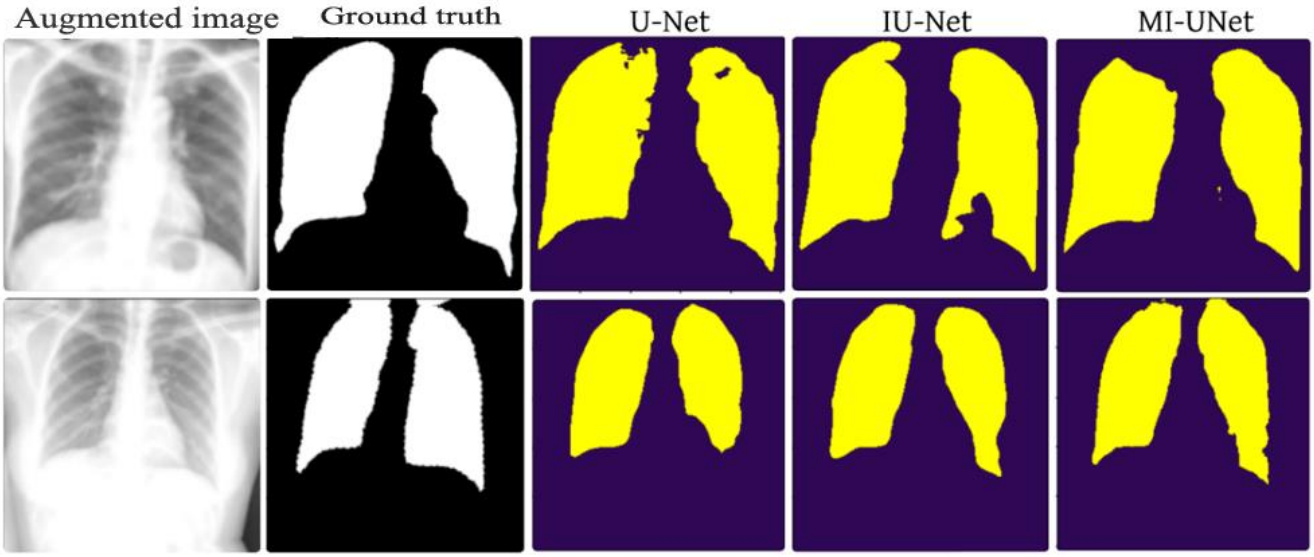


**Figure 15.**
Segmentation results after augmentation with Gaussian blurring.

*5.4. Comparison of Model Training Time Duration*

In addition to analyzing the segmentation performance, an important aspect is to evaluate the time required to train the models. Table 6 compares the training time of the UNet, Unet and MIU-Net models on the original dataset and after augmenting it with CLAHE and Gaussian blurring.

**Table 6.**
Training time comparison.

| Model | Original | CLAHE | Gaussian blurring |
|---|---|---|---|
| U-Net | 3344.12 | 4997.96 | 4801.42 |
| IU-Net | 841.44 | 2402.77 | 2376.05 |
| MIU-Net | 5040.18 | 8401.02 | 8275.18 |

Comparing the training times of the models after augmentation, it can be observed that IU-Net is the fastest model. Despite the increase in training time after augmentation, IU-Net is trained significantly faster compared to U-Net and MIU-Net. For example, after augmentation with Gaussian blur, the training time of IU-Net is 2376.05 seconds, which is about 39.6 minutes, while for U-Net and MIU-Net it is 4801.42 seconds, which is about 80 minutes, and 8275.18 seconds, which is about 137.9 minutes, respectively.

## 6. Discussion

This study proposes the MI-UNet model for the semantic segmentation of biomedical images, specifically focusing on x-ray lung segmentation, advancing the U-Net architecture by integrating Inception and SE blocks. Experiments conducted on the Shenzhen chest X-ray dataset demonstrate that MI-UNet surpasses the performance metrics of the base U-Net model and the modified IU-Net model in terms of Mean IOU, and Dice coefficient. Specifically, for x-ray lung segmentation, MI-UNet achieves a BCE loss of 0.1003, Mean IOU of 0.9040, and Dice of 0.9495, significantly outperforming U-Net (Mean IOU of 0.4493, Dice of 0.6185) and IU-Net (Mean IOU of 0.9035, Dice of 0.9379). Furthermore, incorporating the Inception and Squeeze-Excite blocks into the MI-UNet model provides additional benefits. These blocks allow the model to capture more complex patterns and semantic information from input images, thereby enhancing its ability to accurately segment lung structures on X-rays. The Inception blocks provide multi-scale feature extraction, allowing the model to efficiently capture both local and global contextual information, and the SE blocks facilitate channel-by-channel feature recalibration, highlighting informative features and suppressing irrelevant ones. However, there are limitations to consider. First, the evaluation of MI-UNet effectiveness on a single dataset challenges the model's generalizability. Further testing across a broader range of datasets, especially those including diverse lung diseases such as tuberculosis, pneumonia, and lung cancer, is necessary to confirm the model's versatility and reliability.A second limitation concerns the use of data augmentation methods, such as CLAHE and Gaussian blurring. While these techniques improve the visual quality of images for training, they may not accurately reflect real medical imaging conditions, potentially leading to artificially enhanced model performance without a corresponding improvement in diagnostic accuracy.Therefore, even though MI-UNet showed promise in x-ray lung segmentation, more research is needed to see how well it works in a wider range of clinical situations and diseases. This will allow for a comprehensive evaluation of the model's potential in real-world applications and its ability to aid in the diagnosis and differentiation of lung diseases, contributing to advancements in medical diagnostics. Future research directions should address these challenges in order to bridge the gap between computational model development and clinical application, ultimately contributing to improved medical diagnosis and patient care. The utilization of advanced machine learning techniques, such as deep learning ensembles and transfer

learning, should also be emphasized. These approaches can potentially improve the reliability and accuracy of the MI-UNet model by utilizing different datasets and pre-trained models, thereby increasing its predictive performance and applicability for diagnosing a wider range of lung diseases in different populations and using imaging technologies.

## 7. Conclusion

In the Shenzhen dataset lung X-ray image segmentation study, the IU-Net model performed best with Mean IOU 0.9131 and Dice 0.9361 after Gaussian Blur augmentation (Table 5), while MIU-Net dominated on the original data with Dice 0.9495 (Table 3). The application of CLAHE also improved the results of all models, especially MIU-Net with Dice 0.9449 (Table 4). These results highlight the importance of choosing a magnification method to optimize the accuracy of the segmentation model when studying segmentation of lung X-ray images in the field of healthcare.

## References


[1] K. C. Thandra, A. Barsouk, K. Saginala, J. S. Aluru, and A. Barsouk, "Epidemiology of lung cancer," *Contemporary Oncology*, vol. 25, no. 1, pp. 45-52, 2021.

[2] S. Stirenko *et al.*, "Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation," presented at the 2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO), 2018.

[3] A. S. Panayides *et al.*, "AI in medical imaging informatics: Current challenges and future directions," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, pp. 1837-1857, 2020.

[4] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Systems with Applications*, vol. 138, p. 112821, 2019. https://doi.org/10.1016/j.eswa.2019.112821

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, *Springer International Publishing*, 2015, pp. 234-241.

[6] L. Liu, J. Cheng, Q. Quan, F.-X. Wu, Y.-P. Wang, and J. Wang, "A survey on U-shaped networks in medical image segmentations," *Neurocomputing*, vol. 409, pp. 244-258, 2020. https://doi.org/10.1016/j.neucom.2020.05.070

[7] E. R. C. Q. Brioso, "Anatomical segmentation in automated chest radiography screening," 2022.

[8] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods," *Sustainability*, vol. 13, no. 3, p. 1224, 2021. https://doi.org/10.3390/su13031224

[9] A. Khanna, N. D. Londhe, S. Gupta, and A. Semwal, "A deep residual U-net convolutional neural network for automated lung segmentation in computed tomography images," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1314-1327, 2020. https://doi.org/10.1016/j.bbe.2020.07.007

[10] G. Singadkar, A. Mahajan, M. Thakur, and S. Talbar, "Deep deconvolutional residual network based automatic lung nodule segmentation," *Journal of Digital Imaging*, vol. 33, pp. 678-684, 2020. https://doi.org/10.1007/s10278-019-00301-4

[11] Z. Zhang, C. Wu, S. Coleman, and D. Kerr, "DENSE-INception U-net for medical image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 192, p. 105395, 2020. https://doi.org/10.1016/j.cmpb.2020.105395

[12] A. Mukasheva, D. Koishiyeva, Z. Suimenbayeva, S. Rakhmetulayeva, A. Bolshibayeva, and G. Sadikova, "Comparison evaluation of unet-based models with noise augmentation for breast cancer segmentation on ultrasound images," *Eastern-European Journal of Enterprise Technologies*, vol. 125, no. 9, pp. 85–97, 2023. https://doi.org/10.15587/1729-4061.2023.289044

[13] A. Mukasheva, Z. Akanov, and D. Yedilkhan, "Research of the regression analysis methods for predicting the growth of patients with diabetes mellitus," presented at the 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), IEEE, 2021.

[14] G. Omarova, V. Starovoitov, A. Z. Zh, S. Bekbolatov, A. Ostayeva, and O. Nuridinov, "Application of the Clahe method contrast enhancement of X-Ray images," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, pp. 1-10, 2022. https://doi.org/10.14569/ijacsa.2022.0130549

[15] J. Talukdar, A. Biswas, and S. Gupta, "Data augmentation on synthetic images for transfer learning using deep CNNs," presented at the 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), IEEE, 2018.

[16] D. Yang, G. Liu, M. Ren, B. Xu, and J. Wang, "A multi-scale feature fusion method based on u-net for retinal vessel segmentation," *Entropy*, vol. 22, no. 8, p. 811, 2020. https://doi.org/10.3390/e22080811

[17] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132-7141.

[18] H. Gao, H. Yuan, Z. Wang, and S. Ji, "Pixel transposed convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 5, pp. 1218-1227, 2019.

[19] R. Ballester-Ripoll, P. Lindstrom, and R. Pajarola, "TTHRESH: Tensor compression for multidimensional visual data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 9, pp. 2891-2903, 2019. https://doi.org/10.1109/tvcg.2019.2904063

[20] J. Hu, L. Shen, S. Albanie, G. Sun and E. Wu, "Squeeze-and-Excitation Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 8, pp. 2011-2023, 1 Aug. 2020, doi: 10.1109/TPAMI.2019.2913372.

[21] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv preprint arXiv:1811.03378*, 2018. https://doi.org/10.21275/sr231205140623

[22] P. Murugan and S. Durairaj, "Regularization and optimization strategies in deep convolutional neural network," *arXiv preprint arXiv:1712.04711*, 2017.

[23] F. Kamalov and H. H. Leung, "Deep learning regularization in imbalanced data," presented at the 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), 2020.

[24] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, p. 475, 2014. https://doi.org/10.3390/data7070095

[25] E. Bisong, *Building machine learning and deep learning models on Google cloud platform*. Berkeley, CA: Apress, 2019.

[26] Ange Tato and Roger Nkambou (2018). "Improving Adam Optimizer." Available at:https://openreview.net/forum?id=HJfpZq1DM

[27] B. A. Trinh, T. T. A. Trinh, L. Vu, H. Dao, and T. Nguyen, "Improving loss function for a deep neural network for lesion segmentation," in *Proceedings of the 12th International Symposium on Information and Communication Technology*, 2023, pp. 592-598.

[28] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, pp. 1-28, 2015. https://doi.org/10.1186/s12880-015-0068-x

[29] B. Cheng, R. Girshick, P. Dollár, A. C. Berg, and A. Kirillov, "Boundary IoU: Improving object-centric image segmentation evaluation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15334-15342.

[30] Danilov, Viacheslav; Proutski, Alex; Kirpich, Alexander; Litmanovich, Diana; Gankin, Yuriy (2022). "Chest X-ray dataset for lung segmentation". Mendeley Data, V1. DOI: 10.17632/8gf9vpkhgy.1. Available at: https://data.mendeley.com/datasets/8gf9vpkhgy/1.