

COVID-19 Infection Segmentation on Chest X-ray Images

A U-Net Baseline Using COVID-QU-Ex (COVID-only Training)

Nguyen The Cuong
22BA13058

Machine Learning in Medicine 2026
University of Science and Technology of Hanoi (USTH)
Email: cuongnt.22ba13058@usth.edu.vn

Abstract—This report studies infection region segmentation for COVID-19 chest X-ray images using the COVID-QU-Ex dataset. The goal is to predict a binary mask that highlights the infection area in each input image. Because the dataset is large, we only use the COVID-19 subset for training and evaluation. We implement a U-Net segmentation model and train it with binary cross-entropy loss. We evaluate the model mainly with training curves and Root Mean Squared Error (RMSE) on both training and validation sets. The results show that U-Net can learn meaningful infection patterns, but the segmentation quality is still affected by class imbalance and unclear boundaries in X-ray images.

Index Terms—COVID-19, chest X-ray, segmentation, U-Net, COVID-QU-Ex, RMSE

I. INTRODUCTION

Chest X-ray (CXR) imaging is commonly used to support diagnosis and monitoring for lung diseases, including COVID-19. Many studies focus on classification (COVID vs non-COVID), but segmentation is also important because it gives a visual explanation: the model shows *where* the infection is located.

In this practical work, we build one deep learning model for infection segmentation on chest X-ray images. The task is binary segmentation: each pixel is either infection or background. We use the COVID-QU-Ex dataset and train a U-Net model to predict infection masks.

To reduce the training time and storage cost, we only use the COVID-19 subset (COVID-19 images and their infection masks). This means our model is not trained to handle Normal or Non-COVID images. However, this setting is still useful for learning the infection region patterns among COVID-19 cases.

II. DATASET

We use the COVID-QU-Ex dataset for infection segmentation. Each sample contains a grayscale chest X-ray image and a corresponding infection mask. The mask indicates the infection region using pixel-level labels.

A. Subset Choice (COVID-only)

The full dataset contains several categories (COVID-19, Non-COVID, Normal) and standard splits (Train/Val/Test). Due to the large size of the full dataset and limited computing resources, we choose to train and evaluate using only the **COVID-19 subset**.

B. Preprocessing

All images and masks are resized to 256×256 . Images are normalized to the range $[0, 1]$. Masks are binarized using a threshold of 0.5 (infection vs background).

C. Visual Check

We first visualize random samples to confirm that images and masks are aligned correctly. Figure 1 shows an example of a chest X-ray with the infection mask overlay.

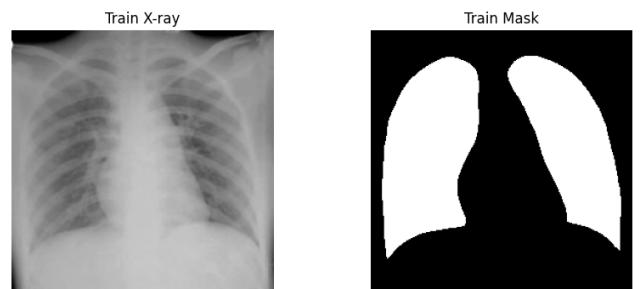


Fig. 1: Example overlay of infection mask on a COVID-19 chest X-ray.

III. METHODS

A. Model Architecture

We use U-Net, a popular encoder-decoder network for medical image segmentation. The encoder extracts features at different scales, and the decoder upsamples the features back to the original size. Skip connections are used to combine low-level spatial information with high-level semantic features.

The model input is a single-channel chest X-ray image, and the output is a probability map (same resolution) representing infection likelihood for each pixel.

B. Training Setup

We train the model using the Adam optimizer with learning rate 10^{-4} and batch size 8. We train for a fixed number of epochs. The loss function is binary cross-entropy.

C. Evaluation Metric: RMSE

In this report, we mainly use RMSE because it is easy to track and compare across training and validation. RMSE is computed between the predicted probability mask and the ground-truth binary mask:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - y_i)^2}, \quad (1)$$

where p_i is the predicted probability for pixel i and y_i is the ground truth label. We report RMSE for training and validation across epochs and also report test RMSE after training.

IV. RESULTS

A. Training Curves

Figure 2 reports the evolution of the training and validation loss over epochs, while Figure 3 shows the corresponding RMSE. Overall, both metrics decrease steadily, indicating that the model continues to learn meaningful representations throughout training. The validation curves closely track the training curves, suggesting stable generalization and no clear evidence of overfitting within the trained epoch range.

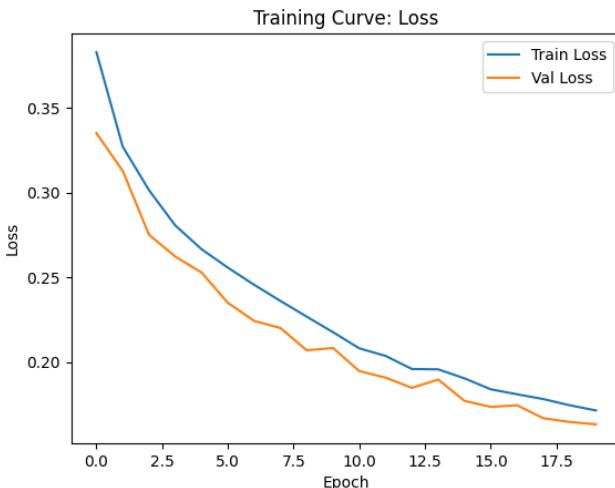


Fig. 2: Training and validation loss across epochs.

A consistent downward trend indicates effective optimization.

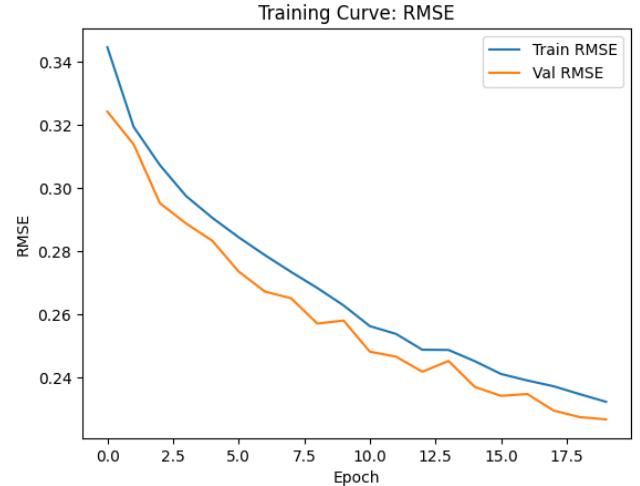


Fig. 3: Training and validation RMSE across epochs.

Both curves decrease and remain close, indicating stable learning and generalization.

B. Test Performance

After training, we evaluate the model on the COVID-19 test split. The final test performance is summarized below:

- **Test RMSE: 0.2288**

C. Qualitative Segmentation Results

To complement quantitative metrics, we visualize representative predictions. Figure 4 shows an example consisting of the input X-ray, the ground-truth mask, and the predicted mask (thresholded at 0.5). The model captures the main infected regions and their coarse shapes, while remaining errors typically appear as small isolated false positives and minor boundary mismatches.

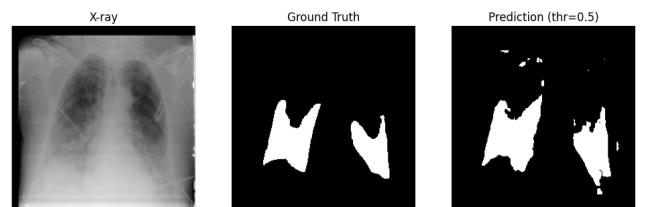


Fig. 4: Qualitative comparison: input X-ray (left), ground-truth mask (middle), and predicted mask (right) at threshold 0.5.

For additional interpretability, Figure 5 overlays the predicted infection region on the original X-ray, making it easier to verify that predicted positives align with plausible lung locations.

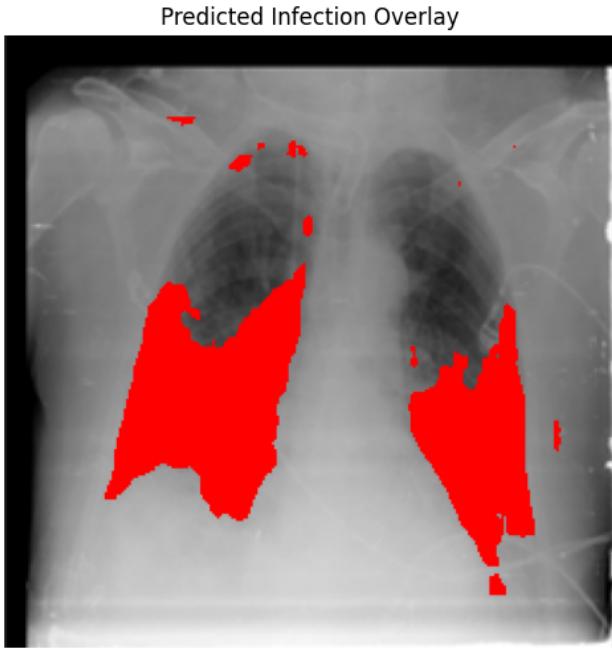


Fig. 5: Predicted infection overlay on the input X-ray (highlighted region indicates predicted infected pixels).

V. DISCUSSION

Overall, the training curves show that the model can learn useful patterns from COVID-19 chest X-rays. RMSE usually decreases during training, which means the predicted mask becomes closer to the ground truth.

However, infection segmentation on chest X-ray is still difficult. First, infection regions often cover a small part of the image, so the background dominates most pixels. This class imbalance can make training unstable or slow. Second, the infection boundary is not always clear, even for humans. Some areas look similar to normal tissue or other structures like ribs. In these cases, the model may miss small infection regions or segment extra regions.

Another limitation is our dataset choice. Since we only train on COVID-19 samples, our model is not designed to segment images from Non-COVID or Normal groups. This decision was made for practical reasons (dataset size and training time), but it reduces the scope of the model.

VI. CONCLUSION AND FUTURE WORK

In this practical work, we implemented a U-Net model for COVID-19 infection segmentation on chest X-ray images using the COVID-QU-Ex dataset. To reduce the dataset size, we trained only on the COVID-19 subset. We evaluated the model mainly using training curves and RMSE on training, validation, and test splits.

For future work, we can improve the project in several ways:

- Add data augmentation (flip, rotation, brightness) to improve generalization.

- Try stronger losses for imbalanced segmentation, such as Dice loss or focal loss.
- Train with more categories (Non-COVID and Normal) if more resources are available.
- Report more segmentation metrics such as Dice and IoU for better comparison.

VII. REFERENCES

- Dataset (COVIDQU): <https://www.kaggle.com/datasets/anasmohammedtahir/covidqu/data>