Neonatal Lung Segmentation in Chest MRI: A Supervised Deep-Learning Approach

Dhyey Patel, Alex Matheson PhD, Abdullah Bdaiwi, PhD, Jason Woods PhD

Cincinnati Children's® changing the outcome together | University of CINCINNATI

Introduction

Lung segmentation in neonatal MRI scans is difficult due to limited data, motion artifacts, and anatomical variability. Manual segmentation is slow and labor-intensive.

Accurate lung segmentation is critical for diagnosing neonatal lung diseases like bronchopulmonary dysplasia. Automating this process saves time and supports faster clinical decisions.

A deep learning model trained specifically on neonatal data can produce fast, accurate lung masks, outperforming adult-trained models.

- Customized **U-Net + ResNet-50** architecture
- Trained on an anonymized neonatal MRI dataset
- Produces masks in seconds, reducing expert workload
- Tailored for **neonatal anatomy**, not adapted from adult datasets

Deep Learning Model

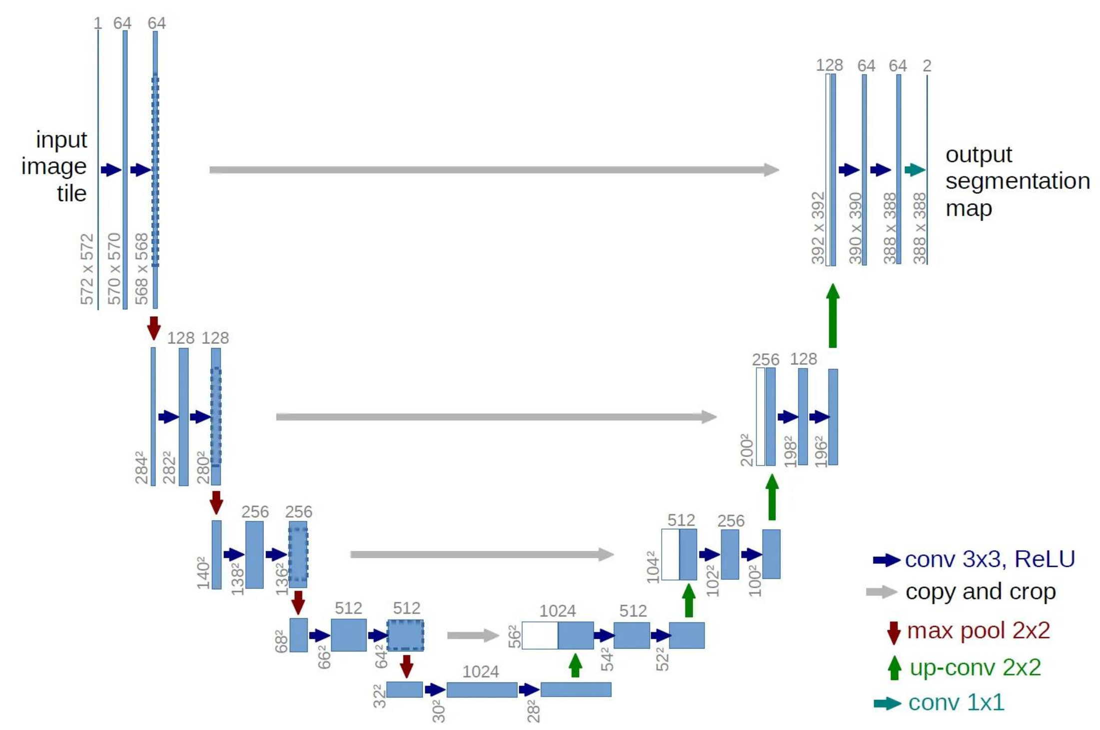

Figure 2 Customized U-Net with ResNet-50 encoder.

Architecture

- **Encoder:** ResNet-50 backbone (ImageNet weights).
- **Decoder:** U-Net structure with skip connections merges multi-scale encoder-decoder features, progressively refining segmentation detail (filters: 256→128→64→32→16).
- Batch normalization stabilizes training and enhances generalization. Sigmoid activation outputs pixel-wise lung segmentation probabilities.

Parameters

- Neonatal MRI dataset (Cohort 1: n=13, Cohort 2: n = 128)
- 70% training, 20% validation, 10% testing
- Combined Binary Cross Entropy and Dice Loss function
- **Optimizer:** Adam (learning rate = 0.001)
- Early stopping, Learning Rate reduction
- Rotations (±15 deg), translations and zoom (±10%).

Limitations and Future Work

Limitations

- Neonatal MRI chest scans are scarce due to ethical and practical constraints, impacting generalizability
- **Slice-wise Segmentation:** Model treats MRI slices independently, ignoring 3D spatial context, potentially causing inconsistencies between slices.
- **Ground Truth Errors:** Inaccurate manual annotations occasionally led to misclassifications, affecting training and validation performance.
- **Artifact Sensitivity:** Motion artifacts and medical equipment may reduce accuracy in certain slices.
- Model performance verified only on CCHMC's CPIR dataset; external validity remains unexplored.

Future Work

- Expand dataset through multicenter collaboration or older infant scans.
- Integrate 3D architectures (e.g., 3D U-Net) to leverage spatial continuity.
- Implement post-processing (morphological filtering, anatomical constraints) to enhance accuracy.
- Validate externally anda develop clinician-focused user interface to refine AI outputs.

Method

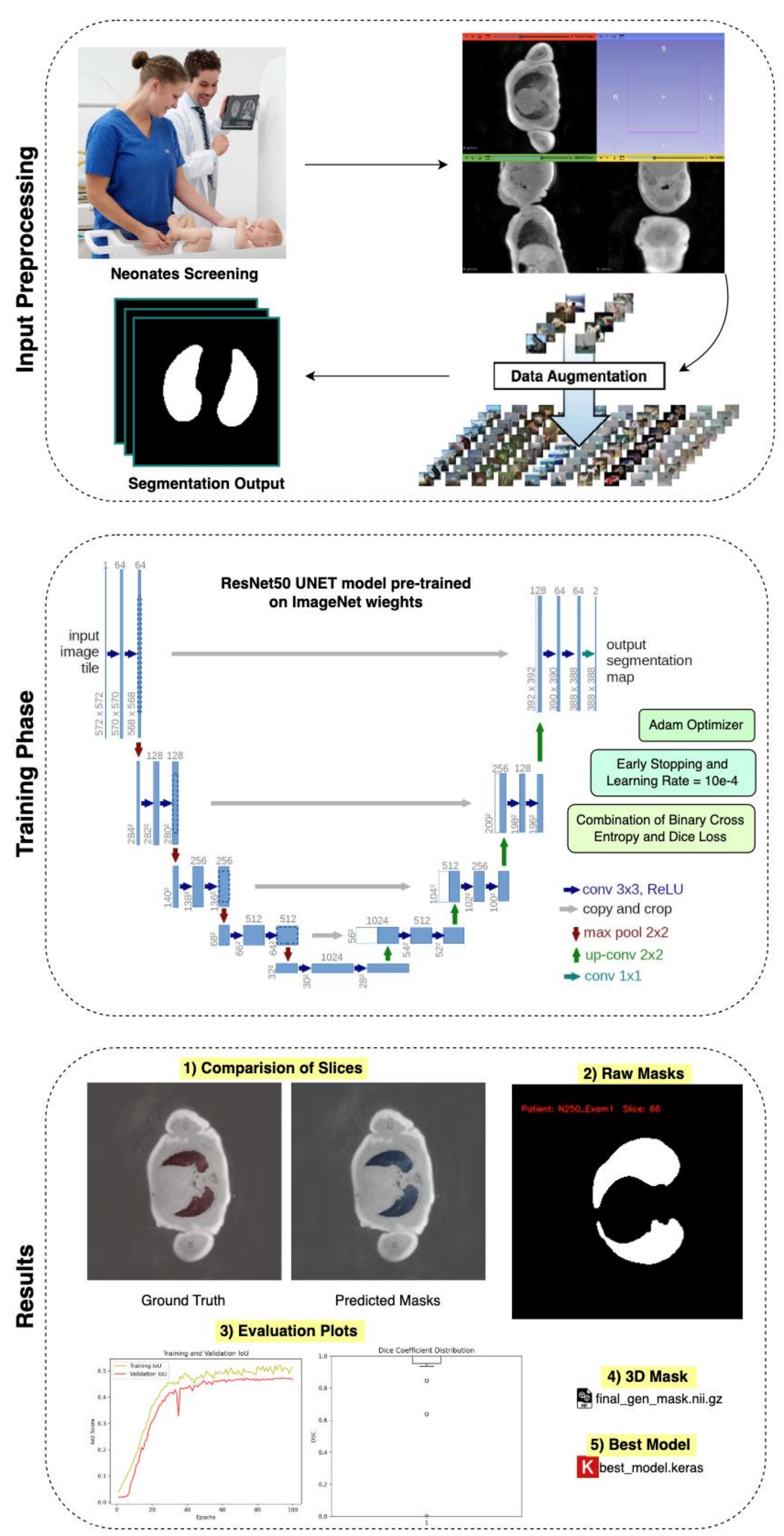

Figure 1 Data Preprocessing, Training, and Results Pipeline

Results

- Mean Dice ≈ **0.90** and IoU ≈ **0.83–0.85** on validation data.
- Slice-level performance: High variability; Dice ranges from **~0.80** (challenging apex/base regions) to **>0.95** (mid-lung slices).
- Clear delineation of lung boundaries: Effectively excludes the heart and other thoracic structures, refer Figure 3.1.
- Typically, near apex/base or around motion artifacts and imaging tubes, refer Figure 3.2.
- Occasional poor predictions due to annotation inconsistencies in **ground truth** as seen in Figure 3.3.
- Generates segmentation masks within seconds, significantly reducing clinical workload.

Ground Truth Overlay | Predicted Mask Overlay

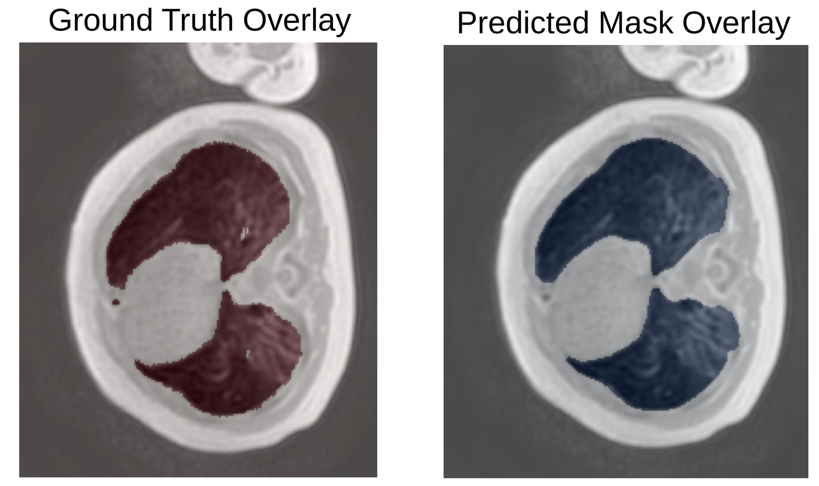

Figure 3.1 Accurate segmentation capturing lung contours clearly and excluding non-lung regions (Dice>0.95).

Ground Truth Overlay | Predicted Mask Overlay

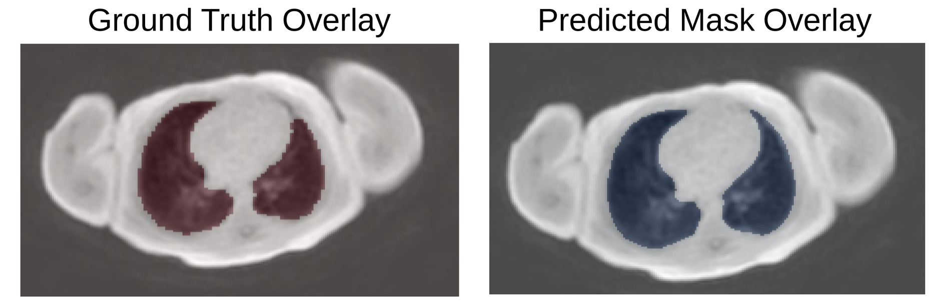

Figure 3.2 Excellent overlap between predicted mask and ground truth, showing robust model generalization.

Ground Truth Overlay | Predicted Mask Overlay

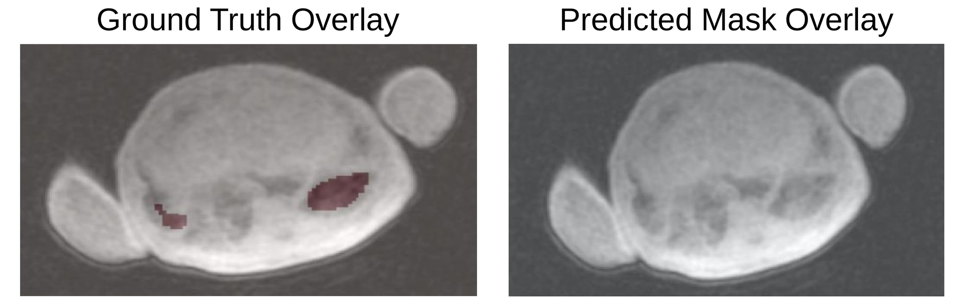

Figure 3.3 Poor annotation leading to a mismatch in ground truth and prediction; plausible model's prediction.

Evaluation Metrics

IoU & Loss Curves (Training vs Validation) (Fig 4.1):

- Steady improvement and convergence, validation loss plateaus after approximately 40 epochs.
- Model shows good generalization without significant overfitting.

Dice Coefficient (Boxplot) (Fig 4.2):

- Median Dice ≈ 0.90 with majority above 0.85.
- Highlights consistent segmentation performance across slices.

ROC Curve (Fig 4.3):

- AUC = 0.999, indicating strong discriminative capability in distinguishing lung and background pixels.

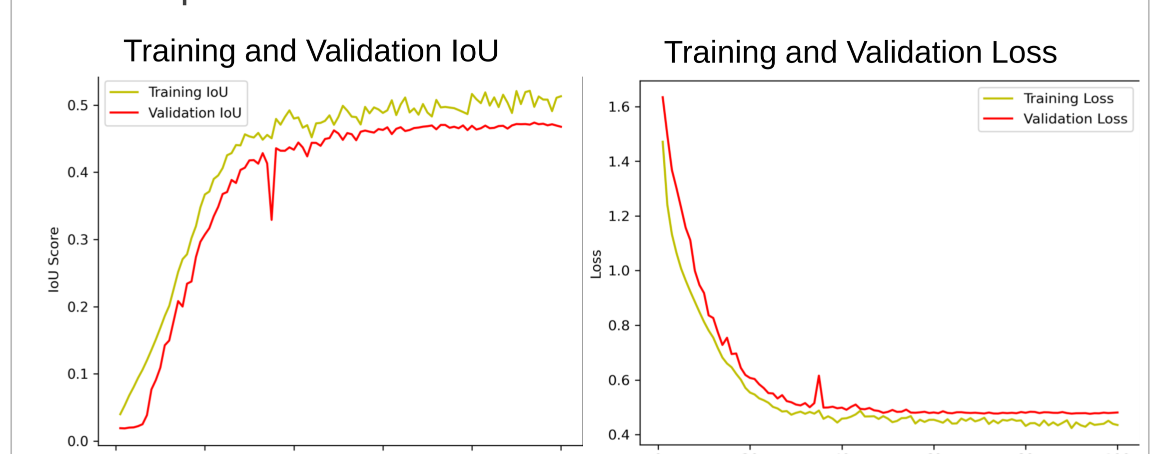

Figure 4.1 Demonstrates successful model training and robust validation performance.

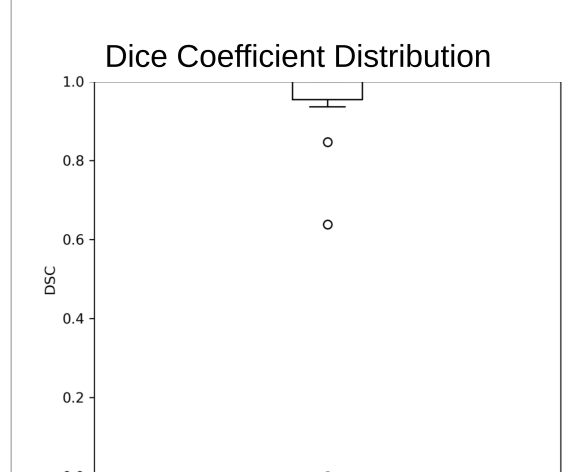

Figure 4.2 Dice distribution indicating reliable performance, with few challenging outlier.

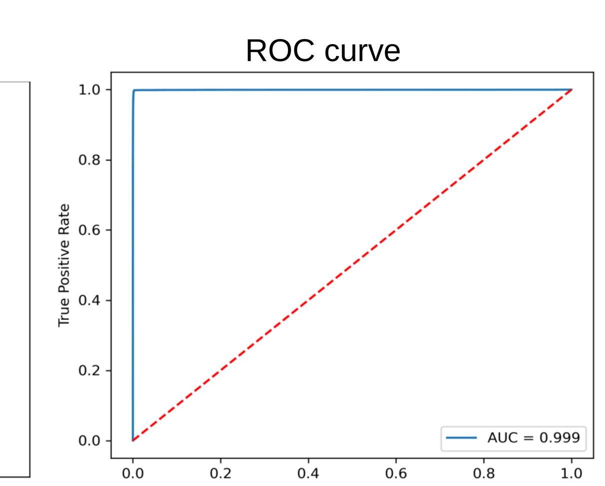

Figure 4.3 Exceptionally high ROC AUC shows the model's strong accuracy in binary lung segmentation.

Conclusion

- Developed a supervised deep-learning model (**U-Net** + **ResNet-50)** for **neonatal lung segmentation in MRI scans**, achieving a strong mean Dice accuracy of ~0.90.
- The model significantly **reduces manual segmentation workload from months to mere seconds**, streamlining clinical workflows despite limited neonatal MRI data, anatomical variability, and motion artifacts.
- Occasional inaccuracies primarily arise from inconsistencies or errors in ground-truth labeling, highlighting the need for reliable annotations.
- Although not yet matching performance benchmarks set by adult lung segmentation models, this neonatal-specific approach dramatically improves efficiency and accuracy for clinical applications, paving the way for faster diagnoses.

References

1. Ronneberger et al. 2015.
2. Litjens et al., 2017
3. Sousa et al. Appl Sci 2022.
4. Thomas et al., 2019
5. Data Augmentation in Deep Learning. insights.daffodilsw.com
6. Lung MRI Segmentation Methods in Preterm Infants. medrxiv.org
7. ResNet50 U-Net Guide. Idiotdeveloper.com
8. https://github.com/qubvel/segmentation_models
9. Çiçek et al., 2016
10. Shorten & Khoshgoftaar, 2019