MAUN Fen Bil. Dergi., 14, 1, 84-94 Araştırma Makalesi/ Research Article
MAUN J. of Sci., 14, 1, 84-94 DOI : 10.18586/msufbd.1877734

Classification of Atelectasis from Chest X-Ray Images Using Deep Learning Approaches

Ömer Faruk GÖK[1], Fatih DOĞANAY[2], Serdar ABUT[3]

[1]Siirt Provincial Directorate of Security, Siirt, Turkiye
[1]Department of Computer Engineering, Faculty of Engineering, Siirt University, Siirt, Turkiye
[2]Department of Research, Mayo Clinic, Jacksonville, FL, USA
[2]Department of Emergency Medicine, Faculty of Medicine, University of Health Sciences, Istanbul Bakirkoy Dr. Sadi Konuk Health Research Center, Istanbul, Turkiye
[3]Department of Artificial Intelligence and Machine Learning, Faculty of Computer and Information Sciences, Kayseri University, Kayseri, Turkiye

✉: serdarabut@gmail.com [1]0009-0000-1710-1631 [2]0000-0003-4720-787X [3]0000-0002-6617-6688

Geliş (Received): 30.01.2026 Düzeltme (Revision): 26.02.2026 Kabul (Accepted): 27.03.2026

ABSTRACT

Atelectasis, the partial or complete collapse of the lung, requires early diagnosis for effective treatment. This study investigates the effects of Region of Interest (ROI) cropping and transfer learning on the automatic classification of atelectasis from chest X-ray images. Four SqueezeNet-based models were developed using two image types (raw and cropped lung regions) and two training strategies (from-scratch and transfer learning). Models I and II were trained from scratch on raw and cropped images, respectively, while Models III and IV employed transfer learning on the same image sets. Performance was evaluated using standard classification metrics. Model IV, combining transfer learning and ROI-cropped images, achieved the best performance, with 98.43% accuracy, 0.984 F1-score, and 0.999 AUC. It also reached 0.9861 precision, 0.9828 sensitivity, and 0.9859 specificity. ROI-based cropping consistently improved classification performance. Combining transfer learning with ROI-focused preprocessing significantly enhances atelectasis detection and shows promise as a reliable clinical decision-support approach.

Keywords: Atelectasis, Deep Learning, Transfer Learning, X-ray Imaging, SqueezeNet, Image Classification

Derin Öğrenme Yaklaşımları Kullanarak Göğüs Röntgeni Görüntülerinden Atelektazinin Sınıflandırılması

ÖZ

Atelektazi, akciğerin kısmen veya tamamen sönmesiyle karakterize edilen ve erken tanı gerektiren önemli bir solunum sistemi hastalığıdır. Bu çalışmada, göğüs röntgeni görüntülerinden atelektazinin otomatik sınıflandırılmasında İlgi Bölgesi (ROI) kırpma ve transfer öğrenmesinin etkileri incelenmiştir. SqueezeNet mimarisi kullanılarak dört farklı model geliştirilmiştir. Modeller, iki görüntü türü (ham ve akciğer bölgesi kırpılmış görüntüler) ile iki eğitim stratejisine (sıfırdan eğitim ve transfer öğrenmesi) göre tasarlanmıştır. Model I ve II sırasıyla ham ve kırpılmış görüntüler üzerinde sıfırdan eğitilmiş, Model III ve IV ise aynı görüntü türlerinde transfer öğrenmesi kullanılarak geliştirilmiştir. Performans standart sınıflandırma ölçütleriyle değerlendirilmiştir. En yüksek performans, transfer öğrenmesi ve ROI kırpma yöntemini birleştiren Model IV tarafından elde edilmiştir. Model %98,43 doğruluk, 0,984 F1 skoru ve 0,999 AUC değeri sağlamıştır. ROI kırpmanın her iki eğitim stratejisinde de performansı artırdığı gözlenmiştir. Transfer öğrenmesi ile ROI tabanlı ön işlemenin birlikte kullanılması, atelektazi tespitinde yüksek başarı sağlayarak klinik karar destek sistemleri için umut verici sonuçlar sunmaktadır.

Anahtar Kelimeler: Atelektazi, Derin Öğrenme, Transfer Öğrenme, X-ışını Görüntüleme, SqueezeNet, Görüntü Sınıflandırması

INTRODUCTION

Atelectasis, defined as the reversible collapse of alveoli due to obstruction of airways, is a common yet complex pulmonary condition that affects normal lung ventilation [1]. While not a disease per se, it frequently appears as a symptom of underlying pulmonary or thoracic disorders. In clinical radiology, it ranks among the most frequently observed abnormalities on chest radiographs and often presents diagnostic challenges [2], partly due to difficulties shared across emergency imaging workflows where AI-assisted diagnostic performance typically ranges between 85–90% in X-ray and CT interpretation [3].

The underlying causes of atelectasis are varied, ranging from airway obstructions caused by inflammation, mucus plugs, or foreign bodies to postoperative complications such as those encountered during anesthesia [4-6]. Despite its diverse etiology, early detection of atelectasis is crucial due to its potential to impair gas exchange and reduce blood oxygenation, thereby affecting overall lung functionality [5]. Chest radiography (CXR) remains the primary diagnostic tool for detecting atelectasis, especially in resource limited settings due to its accessibility and low cost [4, 7]. However, the interpretation of chest X-ray images is not only highly dependent on the expertise of radiologists but also subject to diagnostic ambiguity, particularly in subtle or overlapping cases [2, 8]. These challenges further highlight the broader medical imaging difficulties that fueled the paradigm shift from traditional ANN-based approaches to deep convolutional neural networks (DCNNs) in high dimensional image data analysis [9].

To address these diagnostic difficulties and reduce the subjectivity in image interpretation, artificial intelligence (AI), particularly deep learning (DL), has emerged as a powerful tool in the medical imaging field [9-12]. Deep learning approaches—especially convolutional neural networks (CNNs)—have shown great promise in automating complex visual tasks, including classification, segmentation, and detection of pathologies in radiological images [13-15]. These models can automatically extract hierarchical features from input data [9], offering improved generalizability and reliability over traditional machine learning techniques [14]. Similar to gender recognition studies where feature extraction and preprocessing accuracy significantly affect classification performance [16], medical imaging tasks also benefit from robust, domain-specific feature learning. In medical domains where annotated data are often limited or expensive to collect, transfer learning and model fine-tuning offer further advantages [17, 18], reflectng trends discussed extensively in foundational AI-in-medicine frameworks [12].

The use of deep learning in chest radiograph analysis is particularly notable [19]. Numerous studies have employed CNN architectures such as ResNet, VGG, and DenseNet to detect diseases like pneumonia, tuberculosis, and atelectasis [20-22]. For instance, models such as CheXNet and DLAD-10 have demonstrated performance on par with expert radiologists in detecting a range of thoracic abnormalities from X-ray images [23, 24]. These advancements not only reduce diagnostic workload but also provide fast and consistent preliminary evaluations, which are especially beneficial in emergency settings—an area where AI models have also been proposed for grading diseases such as COPD from chest radiographs [19].

Nevertheless, the performance of deep learning models in medical imaging heavily depends on the quality of preprocessing and the contextual relevance of the input data [11, 15]. Techniques such as contrast enhancement (e.g., CLAHE), image denoising, and lung region segmentation are increasingly employed to improve diagnostic accuracy [25, 26]. Moreover, recent findings suggest that focusing on the anatomical region of interest (ROI)—such as isolating the lung field—can significantly boost model performance [27], which aligns with broader evidence demonstrating the importance of neural network–based feature extraction pipelines in medical decision-making [28]. Deep learning–based anatomical landmark detection and localization methods further enhance this process by reducing spatial uncertainty in structural interpretation [29], while optimized YOLO-based localization frameworks have shown high accuracy in lung region detection through improved anchor box estimation [30].

Despite promising results in the literature, the classification of atelectasis specifically remains underexplored compared to other pulmonary diseases like pneumonia or cardiomegaly [31]. The subtle presentation and variable morphology of atelectasis in chest X-rays make it a challenging target for automated systems. Therefore, there is a critical need for robust and interpretable DL-based classification systems tailored to the early detection of atelectasis. Hybrid learning strategies that combine deep feature extraction with classical classifiers have shown success in related domains [7], suggesting potential value in integrating similar feature-centric approaches into atelectasis classification frameworks.

This study aims to investigate the effectiveness of deep learning models in classifying atelectasis from chest X-ray images. Specifically, it explores how transfer learning and region-focused image preprocessing influence model performance. A total of four classification models were developed using the SqueezeNet architecture—a lightweight convolutional neural network optimized for efficiency. These models were trained on the CheXlocalize[32] dataset, using both original full-frame X-ray images and images cropped to the lung regions.

Two primary research questions guided the study:

- Does transfer learning enhance the classification accuracy compared to training from scratch?
- Does cropping the lung region (i.e., focusing on the ROI) improve diagnostic performance?

The results reveal that models combining transfer learning with ROI-focused preprocessing achieved superior classification performance, reaching 98.43% accuracy, 0.984 F1-score, and 0.999 AUC. These findings demonstrate the high potential of tailored deep learning strategies for detecting atelectasis and underscore the clinical value of incorporating such tools into radiology workflows—consistent with the broader importance of AI-driven decision support in medical imaging [28].

Deep learning has shown promising results in medical image classification, including the detection of lung pathologies. However, several challenges remain, such as ensuring that models focus on medically relevant regions and avoiding overestimation of performance due to data leakage. In this work, we address these

challenges and provide a systematic evaluation of deep learning models for chest image classification.

The main contributions of this study are as follows. First, we demonstrate the importance of careful region-of-interest (ROI) consideration: analysis of Grad-CAM attention maps together with a medical expert revealed that irrelevant regions in X-ray images, such as textual annotations or screen artifacts, can distract model attention. Second, we show that transfer learning allows deep networks to extract useful features efficiently, even with limited training data. Third, unlike many previous studies, we explicitly implement patient-level data splitting to avoid potential data leakage, ensuring that the reported performance reflects the true predictive capability of the models. Overall, this study provides a controlled evaluation framework that highlights how deep learning models focus on medically relevant image regions while minimizing confounding factors such as random splits or leakage.

THE METHOD AND APPLICATION

Dataset

Atelectasis, 720 images from 700 unique patients were initially examined. To ensure data independence, only one image per patient was selected. There were 228 patients with Atelectasis findings and 94 patients labeled as normal. To balance the normal group, 134 additional patients with at least one non-Atelectasis pathology were randomly included, resulting in two groups: 228 images with Atelectasis and 228 images without Atelectasis findings.

Deep Learning Models

This study aimed to build a robust deep learning model for the automatic classification of Atelectasis. Convolutional Neural Networks (CNNs), particularly the SqueezeNet architecture [33], were employed due to their efficiency and effectiveness in image classification tasks. SqueezeNet is known for having a small number of parameters, making it advantageous in distributed computing environments [34] while achieving competitive accuracy levels compared to larger CNNs [35].

The SqueezeNet architecture comprises alternating squeeze and expand layers, grouped into multiple Fire modules. It applies strategies such as using mostly 1×1 filters instead of 3×3, reducing the input channels to 3×3 filters, and delaying downsampling to preserve spatial information [36]. Additionally, it does not use fully connected layers, instead applying global average pooling to compute class scores, which significantly reduces memory and computational demands [37]. The architecture is reported to be 50 times smaller than AlexNet while achieving comparable classification accuracy on ImageNet [38].

Experimental Design

Due to the uncertainty in advance regarding which model setup would yield the best performance, two main strategies were used, each involving two different input configurations: (i) raw chest X-ray images, and (ii) cropped images containing only the lung region.

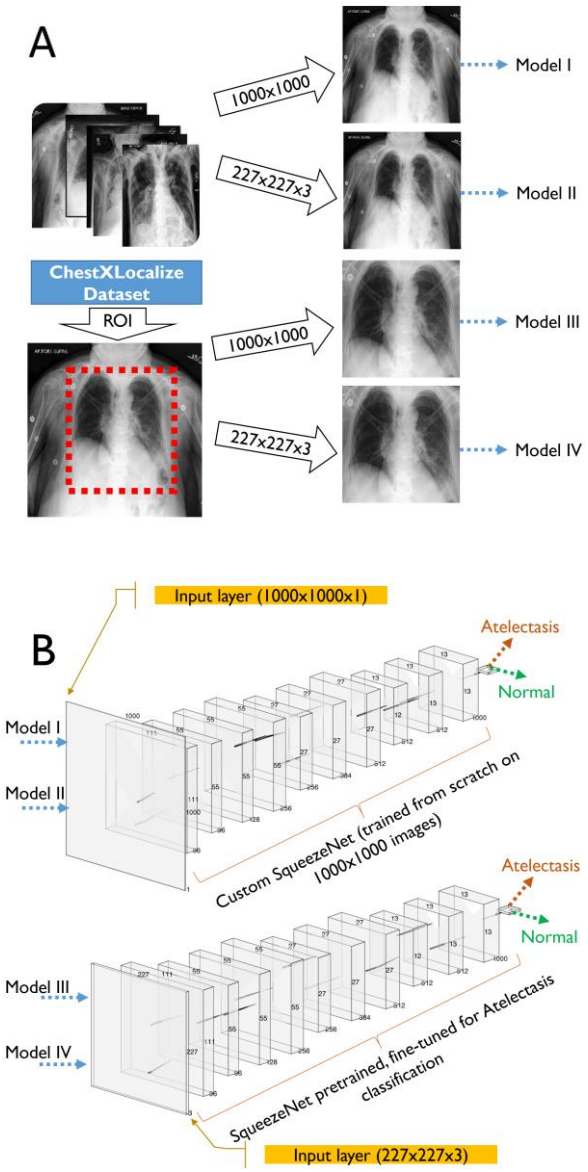


Figure 1. Deep learning model architectures and input preprocessing for Atelectasis classification. Top: SqueezeNet trained from scratch (1000×1000 input). Bottom: pretrained SqueezeNet (227×227×3 input). Both use full and cropped lung images, with outputs classifying Atelectasis vs. Normal.

The first strategy involved training a SqueezeNet model from scratch, with the input size adjusted to 1000×1000 pixels:

Model I: trained on raw full-resolution images

Model II: trained on cropped lung region images, where the cropping process was performed using the method described by [30] to isolate the thoracic area.

This comparison aimed to evaluate the effect of anatomical focus on model performance when training from scratch (Figure 1).
The second strategy used transfer learning with the pretrained SqueezeNet model (input size 227×227×3) originally trained on ImageNet:
Model III: trained on resized raw images
Model IV: trained on resized cropped images, following the same preprocessing pipeline utilized in [30] to ensure consistent lung region segmentation.
This comparison assessed how irrelevant pixel content affects performance in a transfer learning setting (Figure 1).

Training Procedure

To avoid potential data leakage, the dataset was randomly split at the patient level into training (70%), validation (15%), and testing (15%) subsets. Since each image corresponded to a unique patient, patient-level separation was inherently ensured. The dataset was class-balanced (one sample per patient per class); therefore, stratified sampling was not applied. For each model, a separate randomized split was performed, and the reported results are based on a single split (single-fold setting) per model. All models were trained using the Adam optimization algorithm, which adapts learning rates for each parameter using estimates of first and second moments of gradients [39]. Adam combines ideas from AdaGrad and RMSProp and generally requires little hyperparameter tuning [40].
The initial learning rate was set to 0.0001 to balance convergence speed and stability [41]. The mini-batch size was set to 64, a value chosen to ensure efficient computation while maintaining good generalization and memory efficiency [42].
Cross-entropy loss was used as the objective function. It has been shown to be effective in both continuous and combinatorial optimization tasks and helps avoid suboptimal solutions by iteratively updating the sampling distribution [43-45].
The maximum number of training epochs was set to 20. No early stopping criterion was applied, and training was terminated after reaching the predefined epoch limit. For transfer learning, all learnable parameters of the pre-trained networks were updated during training (i.e., no layers were frozen), and a full fine-tuning strategy was employed for all models to allow task-specific feature adaptation.
No data augmentation techniques were applied during training. To ensure a controlled and fair comparison between the evaluated deep learning architectures, the input data pipeline was kept identical across all models without augmentation.

Evaluation Metrics

To evaluate model performance, a set of standard metrics was used, including Accuracy, Precision, Recall (Sensitivity), F1 Score, and ROC-AUC [46-48]. These metrics provide complementary insights into the classifier's ability to distinguish between positive (Atelectasis) and negative cases. Their definitions are as follows:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{3.1}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{3.2}$$

$$Recall(Sensitivity) = \frac{TP}{(TP + FN)} \tag{3.3}$$

$$F1score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{3.4}$$

Additionally, ROC-AUC analysis was employed due to its robustness to class imbalance. ROC curves plot the true positive rate (TPR) against the false positive rate (FPR), providing a comprehensive measure of the classifier's discriminative ability.

$$TPR(Recall) = \frac{TP}{(TP + FN)} \tag{3.5}$$

$$TNR(Specificity) = \frac{TN}{(TN + FP)} \tag{3.6}$$

$$FPR = 1 - TNR = \frac{FP}{(FP + TN)} \tag{3.7}$$

AUC represents the area under the ROC curve, summarizing the model's performance into a single scalar value [49-52].

RESULTS

This section presents a comparative analysis of classification performances of four deep learning models used in this study. The models differ along two main axes: the type of input image (raw vs. cropped) and the training approach (training from scratch vs. transfer learning). Table 1 summarizes the general configuration of the models.

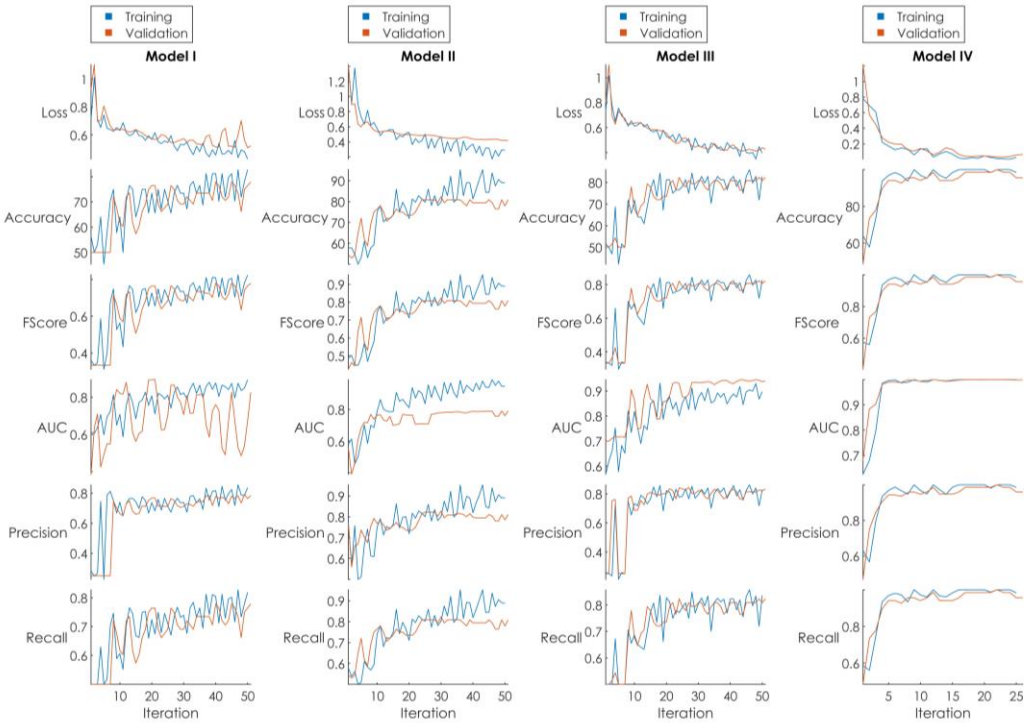


Figure 2. Combined visualization of training and validation curves for accuracy, loss, F1 score, AUC, precision, and recall across all models.

Table 1. Overview of deep learning model configurations: input image type, input size, and training method.

Models	Image Type	Input Size	Training Method
Model I	Raw Images	1000×1000	From scratch
Model II	Cropped (Lung)	1000×1000	From scratch
Model III	Raw Images	227×227×3	Transfer learning
Model IV	Cropped (Lung)	227×227×3	Transfer learning

To compare the four models' performance, metrics such as accuracy, F1 score, area under the curve (AUC), precision, loss, and recall were used. The final training values for these metrics are summarized below

Accuracy Results

As shown in Figure 2, Model I achieved approximately 80% training and validation accuracy by the end of training. Although early epochs showed fluctuations with validation accuracy sometimes lagging behind training accuracy, the gap diminished towards the end, indicating no severe overfitting. Model II demonstrated more stable training with smoother validation accuracy curves, consistently close to the training accuracy from early stages. Ultimately, Model II outperformed Model I in accuracy (~89%) and exhibited less variance.

While Model II presented a steadier validation curve, Model I generally reached higher accuracy peaks, suggesting that high-resolution raw images may enhance generalization when training from scratch. However, Model I's fluctuating validation accuracy hints at occasional overfitting tendencies.

Models III and IV, both based on pretrained SqueezeNet architectures, showed a clear advantage for Model IV, which was trained on cropped lung images. As seen in Figure 2, Model IV's training was more stable, with closely aligned training and validation accuracy curves and minimal variance. This result indicates that combining transfer learning with focused input regions substantially improves classification performance. Cropping to the region of interest enhances both overall accuracy and generalization. Transfer learning is particularly effective in medical imaging tasks with limited data. The best validation accuracy was achieved by Model IV.

Loss Results

Figure 2 shows the loss trends during training and validation phases. Model IV attained the lowest and most stable loss values, achieving rapid decreases early in training and maintaining minimal loss levels. In contrast, Models I and II exhibited higher and more fluctuating loss values, especially in the from-scratch training setting, showing signs of overfitting. These results highlight transfer learning's positive effect on model stability and the benefits of using cropped lung regions for easier model learning.

F1 Score Results

As seen in Figure 2, Model IV outperformed all others in terms of F1 score, maintaining steadily increasing values across both training and validation sets. Model I showed comparatively low F1 scores with significant fluctuations, reflecting poorer classification consistency. This demonstrates that transfer learning combined with lung-region cropping significantly enhances classification accuracy. Since F1 score balances precision and recall, its improvement reflects overall model success.

AUC Results

Figure 2 presents the AUC metrics. Model IV achieved the highest and most consistent AUC values during training and validation, surpassing from-scratch Models I and II, which exhibited volatile validation AUC. The early convergence and high accuracy in Model IV highlight transfer learning's contribution to reliable classification with limited data and the advantage of focusing on the relevant anatomical region.

Precision Results

Precision trends are displayed in Figure 2. Model IV showed the highest precision on the validation set. Similar to Model III, it benefited from transfer learning and the cropped lung region input, reducing false positive predictions. Models I and II exhibited fluctuating and lower precision values during training and validation, underscoring the importance of both transfer learning and input focus in improving decision accuracy.

Recall Results

Recall metrics, shown in Figure 2, further confirm Model IV's superior performance, maintaining the highest recall values throughout training compared to other models. Model III also performed well, benefiting from transfer learning.

Confusion Matrices

The classification performance of four different deep learning models for Atelectasis detection was comparatively evaluated using confusion matrices (Figure 3). Model I, representing a SqueezeNet architecture trained from scratch on full chest X-ray images resized to 1000×1000 pixels, achieved a sensitivity of 94.1% and specificity of 73.5%. Model II, also trained from scratch but using cropped lung regions, showed a slightly reduced sensitivity of 91.2%, while maintaining the same specificity (73.5%) as Model I.

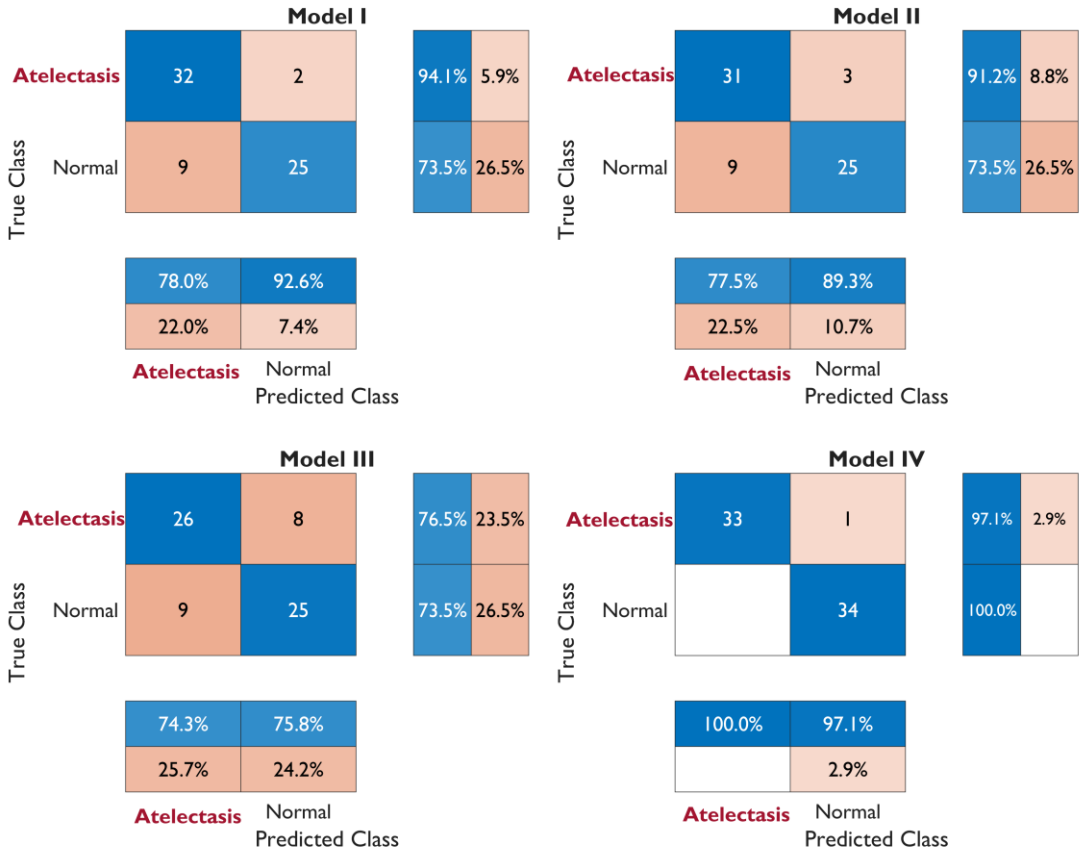


Figure 3. Confusion matrices of four different SqueezeNet-based models used for Atelectasis classification.

Model III, which utilized a pretrained SqueezeNet fine-tuned on full images (227×227 input size), exhibited decreased sensitivity (76.5%) and similar specificity (73.5%), indicating that transfer learning alone without region-focused preprocessing may not sufficiently capture pathological features. Model IV, which combined transfer learning with cropped lung regions, achieved the highest performance, with 97.1% sensitivity and 100% specificity, correctly identifying all Normal cases and nearly all Atelectasis cases.

All models were evaluated on the same test set (N = 68; 34 Atelectasis, 34 Normal), and no data augmentation was used during inference. The results suggest that combining region-focused input (cropped lungs) with pretrained models significantly enhances classification accuracy, particularly by reducing false positives in the Normal class. These performance gains were consistent across metrics and visually supported by the confusion matrices, indicating statistical robustness of the proposed approach.

Grad-CAM Localization Findings of Disease Labels

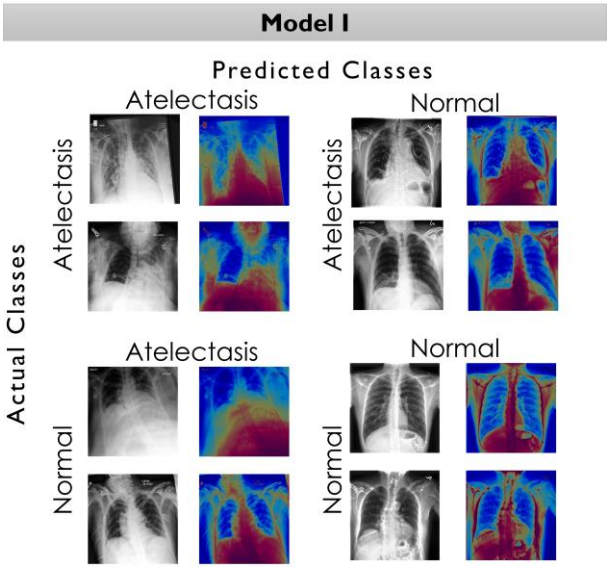


Figure 4. Examples with the highest classification scores in each cell obtained by Model I.

Figure 4 presents two examples with the highest classification scores for each cell from the classification results of Model I (raw image + training from scratch).

The horizontal axis represents the predicted classes ("Atelectasis" and "Normal"), while the vertical axis represents the actual classes. In each example, the left image shows the original chest radiograph, and the right image displays the Grad-CAM-based heatmap.

The top-left cell represents True Positive (TP) examples, the top-right cell False Negative (FN), the bottom-left cell False Positive (FP), and the bottom-right cell True Negative (TN) examples. Although Model I partially identifies pathological regions correctly, its attention distribution occasionally drifts outside the lung area, and both FP and FN rates remain relatively high. This indicates sensitivity to data noise and limited clinical reliability.

Figure 5 presents high-score examples from the classification results of Model II (cropped lung image + training from scratch). By cropping, background noise was reduced, leading to fewer FP cases compared to Model I. In TP examples, the model achieved clearer focus on pathological regions. Although FN cases decreased, misses still occurred in low-contrast or localized lesions. FP cases were limited and generally caused by intensity variations within the lung region.

Figure 6 shows high-score examples from the classification results of Model III (raw image + transfer learning). Thanks to transfer learning, more stable and consistent results were obtained compared to Model I. In TP examples, pathological regions were successfully identified; however, the attention occasionally shifted outside the lung area.

FP cases mostly originated from device artifacts and radiograph edges. The FN rate remained moderate, and detection performance was limited in diffuse lesions.

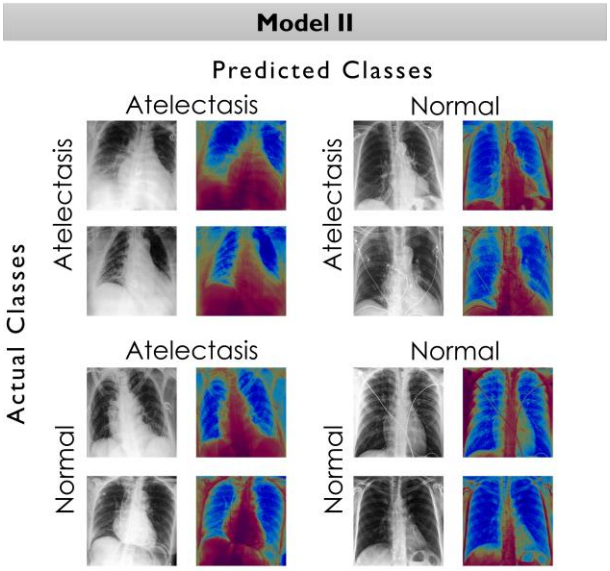


Figure 5. Examples with the highest classification scores in each cell obtained by Model II.

Figure 7 displays the highest-score examples from Model IV (cropped lung image + transfer learning). In TP cases, the model achieved clear and clinically meaningful focus on pathological areas. The FN rate was very low, and no FP cases were observed. In TN examples, low-level and balanced activations enabled a clear separation. This model appears to be the most promising candidate for clinical application, demonstrating both high sensitivity and specificity.

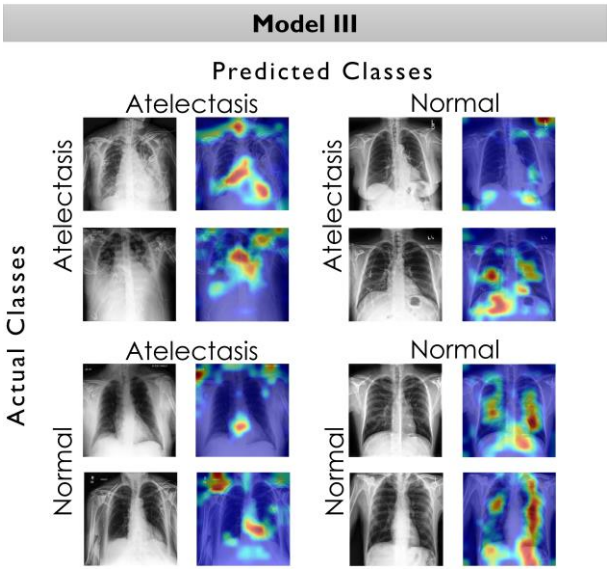
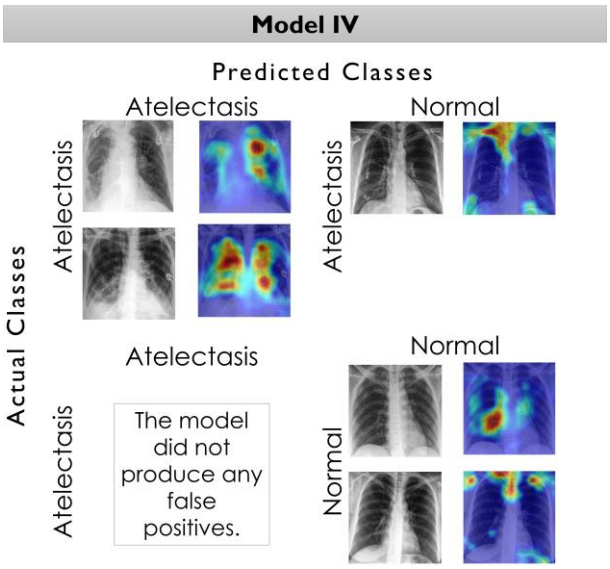


Figure 6. Examples with the highest classification scores in each cell obtained by Model III.



Şekil 7. Examples with the highest classification scores in each cell obtained by Model IV.

Visual Attention Analysis in Misclassifications

Grad-CAM-based heatmaps were used to examine the regions of visual focus that influenced the model's decision-making process. The analysis concentrated on two main types of errors:

- False Negative (FN): True atelectasis cases incorrectly classified as normal.
- False Positive (FP): True normal cases incorrectly classified as atelectasis.

Model I – Raw Image + Training from Scratch

False Negatives (FN): The model's attention was often concentrated outside the lung area, such as the edges of the radiograph, shoulder line, or chest frame. Weak activation was observed in pathological regions, especially in low-contrast or small lesions, indicating dispersed attention. This suggests that training from scratch with limited data was insufficient for learning pathology-specific features.

False Positives (FP): The model frequently misinterpreted edge artifacts, intensity variations, medical device traces (tubes, cables), and vascular densities near the lung borders as pathological findings. Such spurious activations reflect misleading associations, often referred to as spurious correlations.

Model II – Cropped Lung Image + Training from Scratch

False Negatives (FN): Although the model's attention was mostly confined within the lung area, it tended to focus on homogeneous regions or the healthy side of the opposite lung instead of the actual pathology. Localized, non-diffuse, and low-contrast lesions were frequently overlooked.

False Positives (FP): Cropping effectively reduced the influence of edge artifacts. However, vascular shadows, mild opacities, and normal anatomical variations within the lung were sometimes mistaken for pathology.

Model III – Raw Image + Transfer Learning

False Negatives (FN): The model's focus occasionally shifted toward the lung boundaries or mediastinal region. While partial attention was observed on pathological areas, activation intensity was often insufficient, resulting in low sensitivity for diffuse or small lesions. Background noise partially reduced the benefit of transfer learning.

False Positives (FP): Non-pathological structures such as device artifacts, spinal shadows, and vascular density in the upper lobes were often misinterpreted as abnormal. Brightness and intensity variations in these areas led to false alarms.

Model IV – Cropped Lung Image + Transfer Learning

False Negatives (FN): Only a few FN cases were observed. In these instances, the attention was generally close to the pathology but occasionally included secondary focus on the normal lung region. This behavior typically occurred when the lesion was very small or had low contrast.

False Positives (FP): No false positive cases were identified. The complete removal of background noise through cropping, combined with the strong feature extraction capability provided by transfer learning, was decisive in achieving this performance.

CONCLUSION

In this study, the classification performance of four distinct deep learning models for detecting atelectasis was comparatively evaluated. The models differed in terms of input image preprocessing (raw vs. cropped) and training strategy (from-scratch vs. transfer learning). According to the obtained results, Model IV, which utilized transfer learning and was trained on cropped lung region images, achieved the highest and most consistent performance across all metrics (Loss, F-score, AUC, Precision, Recall). In contrast, Model I, which was trained from scratch using raw images, yielded the lowest performance.

These findings demonstrate that transfer learning enables more efficient learning with limited data and that extracting the relevant anatomical region significantly enhances classification performance. Especially in reducing false positives and improving sensitivity, focused modeling of the region of interest proved to be critical. Overall, efficient use of training data and targeting clinically meaningful regions significantly improved deep learning-based classification performance. We note that the test set consisted of only 68 samples (34 per class), which limits the statistical power of the reported performance metrics. High values, such as 100% specificity observed for Model IV, should therefore be interpreted with caution. Larger independent datasets are needed to confirm and generalize these findings.

Table 2. Performance metrics of all models at the end of training.

Model	Acc. (%)	F-score	AUC	Prec.	Loss	Recall
Model I	82.81	0.82	0.89	0.83	0.42	0.81
Model II	89.06	0.88	0.94	0.89	0.28	0.88
Model III	82.81	0.82	0.89	0.82	0.39	0.83
Model IV	**98.43**	**0.98**	**0.99**	**0.98**	**0.024**	**0.98**

Future Work

For future studies, it is recommended to replace manual cropping of the lung region with automated methods. In this regard, segmenting the lungs from the chest X-ray and feeding only the relevant region to the classification models would enhance both overall performance and clinical applicability.

Among the deep learning-based segmentation approaches that can be considered for this purpose are U-Net [53], Mask R-CNN [54], DeepLabV3+ [55], and Attention U-Net [56]. These methods allow for precise segmentation of anatomical structures such as the lungs, thereby providing cleaner and more meaningful input to classification models. Moreover, automating this process reduces preprocessing time and increases the potential for real-time deployment of the models.

ACKNOWLEDGEMENTS

- Funding: F.D. acknowledges a Postdoctoral grant from the Scientific and Technical Research Council of Turkey (TUBITAK, 2219 - International Postdoctoral Research Scholarship Programme, 1059B192302217).
- Ethical Approval: This article complies with ethical standards.

REFERENCES

[1] Peroni, D.G. and A.L. Boner, Atelectasis: mechanisms, diagnosis and management. Paediatric Respiratory Reviews, 2000. 1(3): p. 274-278.

[2] Woodring, J.H. and J.C. Reed, Types and Mechanisms of Pulmonary Atelectasis. Journal of Thoracic Imaging, 1996. 11(2): p. 92-108.

[3] Berikol, G.B., et al., Mapping artificial intelligence models in emergency medicine: A scoping review on artificial intelligence performance in emergency care and education. Turkish Journal of Emergency Medicine, 2025. 25(2): p. 67-91.

[4] Carlsen, K.H., S. Crowley, and B. Smevik, 70 - Atelectasis, in Kendig's Disorders of the Respiratory Tract in Children (Ninth Edition), R.W. Wilmott, et al., Editors. 2019, Elsevier: Philadelphia. p. 1027-1033.e1.

[5] Restrepo, R.D. and J. Braverman, Current challenges in the recognition, prevention and treatment of perioperative pulmonary atelectasis. Expert Review of Respiratory Medicine, 2015. 9(1): p. 97-107.

[6] Hedenstierna, G. and H.U. Rothen, Atelectasis Formation During Anesthesia: Causes and Measures to Prevent It. Journal of Clinical Monitoring and Computing, 2000. 16(5): p. 329-335.

[7] Sajed, S., et al., The effectiveness of deep learning vs. traditional methods for lung disease diagnosis using chest X-ray images: A systematic review. Applied Soft Computing, 2023. 147: p. 110817.

[8] Ayalew, A.M., et al., Atelectasis detection in chest X-ray images using convolutional neural networks and transfer learning with anisotropic diffusion filter. Informatics in Medicine Unlocked, 2024. 45: p. 101448.

[9] Abut, S., H. Okut, and K.J. Kallail, Paradigm shift from Artificial Neural Networks (ANNs) to deep Convolutional Neural Networks (DCNNs) in the field of medical image processing. Expert Systems with Applications, 2024. 244: p. 122983.

[10] Hwang, E.J., et al., Deep Learning for Detection of Pulmonary Metastasis on Chest Radiographs. Radiology, 2021. 301(2): p. 455-463.

[11] Moses, D.A., Deep learning applied to automatic disease detection using chest X-rays. Journal of Medical Imaging and Radiation Oncology, 2021. 65(5): p. 498-517.

[12] Abut, S., et al., Deep Neural Networks and Applications in Medical Research, in Deep Learning - Recent Findings and Research, M. Domínguez-Morales, et al., Editors. 2023, IntechOpen: London.

[13] Yang, S., et al., Intelligent health care: Applications of deep learning in computational medicine. Frontiers in Genetics, 2021. 12: p. 607471.

[14] Ching, T., et al., Opportunities and obstacles for deep learning in biology and medicine. Journal of The Royal Society Interface, 2018. 15(141): p. 20170387.

[15] Bakator, M. and D. Radosav, Deep Learning and Medical Diagnosis: A Review of Literature. Multimodal Technologies and Interaction, 2018. 2(3): p. 47.

[16] Ergen, B. and S. Abut. Gender recognition using facial images.

[17] Jasmine Pemeena Priyadarsini, M., et al., Lung Diseases Detection Using Various Deep Learning Algorithms. Journal of Healthcare Engineering, 2023. 2023(1): p. 3563696.

[18] Kido, S., Y. Hirano, and N. Hashimoto. Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN). in 2018 International Workshop on Advanced Image Technology (IWAIT). 2018.

[19] Abut, S., AI-Based Model Design for Prediction of COPD Grade from Chest X-Ray Images: A Model Proposal (COPD-GradeNet). Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, 2024. 39(2): p. 325-338.

[20] Aggarwal, R., et al., Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. npj Digital Medicine, 2021. 4(1): p. 65.

[21] Geng, Y., et al. 2.5 D CNN model for detecting lung disease using weak supervision. in Medical imaging 2019: computer-aided diagnosis. 2019. SPIE.

[22] Tharakan, J.J., M. Kanchan, and O.S. Powar. Enhancing Disease Detection with Machine and Deep Learning: Analyzing Chest X-rays for Pneumonia and Atelectasis Identification. in 2024 International Conference on Augmented Reality, Intelligent Systems, and Industrial Automation (ARIIA). 2024.

[23] Nam, J.G., et al., Development and validation of a deep learning algorithm detecting 10 common abnormalities on chest radiographs. European Respiratory Journal, 2021. 57(5).

[24] Rajpurkar, P., et al., Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225, 2017.

[25] Chutia, U., A.S. Tewari, and J.P. Singh, Collapsed lung disease classification by coupling denoising algorithms and deep learning techniques. Network Modeling Analysis in Health Informatics and Bioinformatics, 2023. 13(1): p. 1.

[26] Chutia, U., et al., Classification of lung diseases using an attention-based modified DenseNet model. Journal of Imaging Informatics in Medicine, 2024. 37(4): p. 1625-1641.

[27] Chutia, U., A.S. Tewari, and J.P. Singh, Lung disease classification using deep learning and genetic algorithm. Neural Computing and Applications, 2024.

[28] Abut, S.O., Hayrettin, The Importance of Artificial Neural Networks in Decision Making for the Field of Medicine, in The Future of Artificial Neural Networks, I.M. Ghosal, Arun; Jain, Hemlata, Editor. 2024, Nova Science and Technology: New York. p. 1-24.

[29] Oflas, R.D., Fatih; Abut, Serdar. Anatomical Landmark Detection in Medical Imaging: Deep Learning–Based Approaches and Clinical Reliability in 4th International Conference on Pioneer and Innovative Studies ICPIS 2026. 2026. Konya: All Sciences Academy.

[30] Abut, S., IoU-Based Anchor Box Estimation for Enhanced Lung Region Localization in Chest X-rays Using YOLO v4. Kirklareli University Journal of Engineering and Science, 2025. 11(2): p. 253-266.

[31] Chaudhary, A., A. Hazra, and P. Chaudhary. Diagnosis of chest diseases in x-ray images using deep convolutional neural network. in 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). 2019. IEEE.

[32] Saporta, A., et al., Benchmarking saliency methods for chest X-ray interpretation. Nature Machine Intelligence, 2022. 4(10): p. 867-878.

[33] Iandola, F.N., et al., SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360, 2016.

[34] Tsivgoulis, M., T. Papastergiou, and V. Megalooikonomou, An improved SqueezeNet model for the diagnosis of lung cancer in CT scans. Machine Learning with Applications, 2022. 10: p. 100399.

[35] Hassanpour, M. and H. Malek. Document Image Classification using SqueezeNet Convolutional Neural Network. in 2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS). 2019.

[36] Ullah, A., et al., Comparative analysis of AlexNet, ResNet18 and SqueezeNet with diverse modification and arduous implementation. Arabian journal for science and engineering, 2022. 47(2): p. 2397-2417.

[37] Pradeep, K., et al. EdgeNet: SqueezeNet like Convolution Neural Network on Embedded FPGA. in 2018 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS). 2018.

[38] Yuan, C., et al., An effective image classification method for shallow densely connected convolution networks through squeezing and splitting techniques. Applied Intelligence, 2019. 49(10): p. 3570-3586.

[39] Kingma, D.P. and J. Ba, Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[40] Hospodarskyy, O., et al., Understanding the adam optimization algorithm in machine learning. 2024, 2024.

[41] Smaida, M., S. Yaroshchak, and A.B. Sasi, Learning rate optimization in CNN for accurate ophthalmic classification. Int. J. Innov. Technol. Explor. Eng.(IJITEE), 2021. 10: p. 211-216.

[42] Masters, D. and C. Luschi, Revisiting small batch training for deep neural networks. arXiv preprint arXiv:1804.07612, 2018.

[43] Rubinstein, R., The cross-entropy method for combinatorial and continuous optimization. Methodology and computing in applied probability, 1999. 1(2): p. 127-190.

[44] Botev, Z.I., et al., The cross-entropy method for optimization, Handbook of Statistics. Handbook of Statistics. Machine Learning, 2013. 31: p. 19-34.

[45] De Boer, P.-T., et al., A tutorial on the cross-entropy method. Annals of operations research, 2005. 134(1): p. 19-67.

[46] Sharma, A., Testing of machine learning algorithms and models. 2023, Universität Oldenburg.

[47] Botchkarev, A., A new typology design of performance metrics to measure errors in machine learning regression algorithms. Interdisciplinary Journal of Information, Knowledge, and Management, 2019. 14: p. 045-076.

[48] Naidu, G., T. Zuva, and E.M. Sibanda. A Review of Evaluation Metrics in Machine Learning Algorithms. in Artificial Intelligence Application in Networks and Systems. 2023. Cham: Springer International Publishing.

[49] Wu, S. and P. Flach. A scored AUC metric for classifier evaluation and selection. in Second workshop on ROC analysis in ML, bonn, Germany. 2005.

[50] Zhang, X., et al., The use of ROC and AUC in the validation of objective image fusion evaluation metrics. Signal processing, 2015. 115: p. 38-48.

[51] Muschelli III, J., ROC and AUC with a binary predictor: a potentially misleading metric. Journal of classification, 2020. 37(3): p. 696-708.

[52] Marzban, C., The ROC curve and the area under it as performance measures. Weather and Forecasting, 2004. 19(6): p. 1106-1114.

[53] Ronneberger, O., P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. in International Conference on Medical image computing and computer-assisted intervention. 2015. Springer.

[54] He, K., et al. Mask r-cnn. in Proceedings of the IEEE international conference on computer vision. 2017.

[55] Chen, L.-C., et al., Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587, 2017.

[56] Oktay, O., et al., Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999, 2018.