Leveraging Deep Learning Artificial Intelligence to Conduct Quality Control on Chest X-ray Images

M Zhang[1] , K Nye[2] , G Avinash[1] , JM Sabol[2]*,

[1]GE Healthcare, San Ramon, CA, [2]GE Healthcare, Waukesha, WI

ABSTRACT

Purpose: To develop and assess performance of artificial intelligence (AI) algorithms for automatic quality control (QC) checks on chest x-ray (CXR) images.

Methods: Over 100,000 x-ray images of numerous anatomical exams and views were compliantly collected from five institutions in the USA, Canada, and China. Variational autoencoder (VAE) deep learning and convolutional neural networks (CNNs) were used to create AI algorithms to automatically perform two QC tasks. For the first task, a one-class classifier using VAE and CNNs, was trained to detect an adult frontal (AP/PA) CXR image versus image of other anatomy. For the second task, a binary classifier using CNNs, was trained to determine whether the patient positioning in an adult frontal CXR was acceptable. Performance of the algorithms for correct view and positioning were evaluated using receiver operating characteristic (ROC) analysis.

Results: Both algorithms performed very effectively; each with an ROC area-under-the-curve of 0.99. The accuracies of the algorithms were 0.99 and 0.95 for the frontal CXR detection algorithm and patient positioning algorithm respectively.

Conclusion: This work demonstrates the feasibility of using AI to determine if incorrect anatomy or view were acquired and whether patient positioning was acceptable for CXR images. These results warrant further development to expand anatomy and view types, and additional image reject reasons.

PURPOSE

Continuous Quality Control (QC) in x-ray imaging is vital for a high quality radiology department. Effective QC efforts are a continuous process involving time-consuming effort of the medical physicist for collection of data from multiple imaging systems and laborious analysis. Repeated and rejected x-ray images result in unnecessary radiation exposure to the patient and inefficiency in the radiology department, delaying rejected x-ray images is a key component of a successful QC program. We propose that deep learning (DL) algorithms can minimize the effort and improve accuracy of QC programs. In this poster, we describe the development and performance assessment of DL algorithms that perform automatic QC checks on CXR images. We focus on (1) if the acquisition protocol used appropriately matched a frontal CXR image acquired and (2) was the frontal CXR positioning acceptable?

TASK 1: DETECT FRONTAL CXR VIEW

Data Sources and Number of Images

Variational AutoEncoder (VAE): USA Hospital 2 (90,000 for training, 20,000 for validation)

Convolutional Neural Network (CNN): Canada, USA Hospital 1: 13,150 (non-chest frontal), 1,892 (chest-frontal), China, USA Hospitals 2 & 3: 11,823 (chest-frontal)

Ratio of Training : Validation : Testing Cases: 60% : 20% : 20%

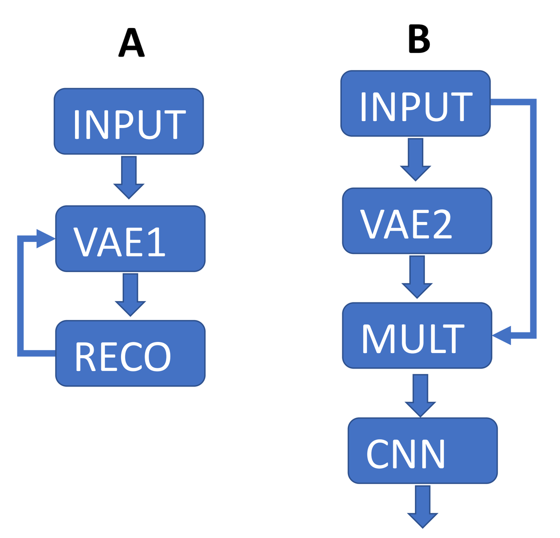
Training:

1. Data Augmentation (each epoch: each batch) randomly pick *N* (*N*=batch size) images from the training set, apply the rotation, shift, shear, zoom, horizontal flips
2. There would be augmented *number of batches * N* data instances for each epoch to train the model.
3. Keep the model that performed best on the augmented validation data as the best model thus far.

Key Points:

1. The use of a VAE + CNN Network is a novel approach
2. Serves as base model for other CXR frontal-view applications (gate model for filtering the data)
3. Accurate and robust methodology

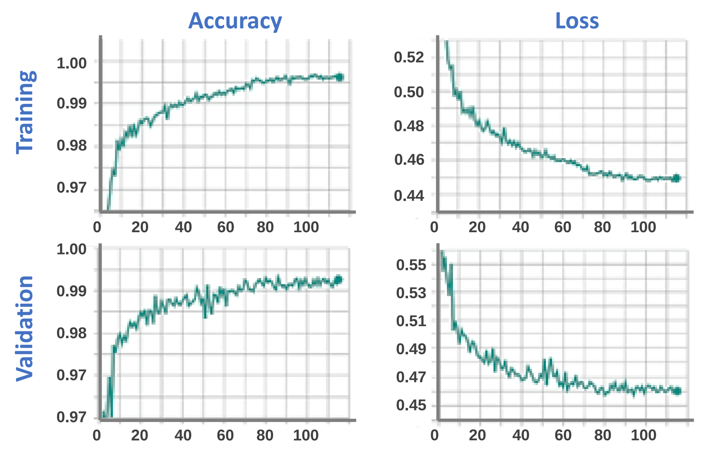
Network (VAE + CNN)



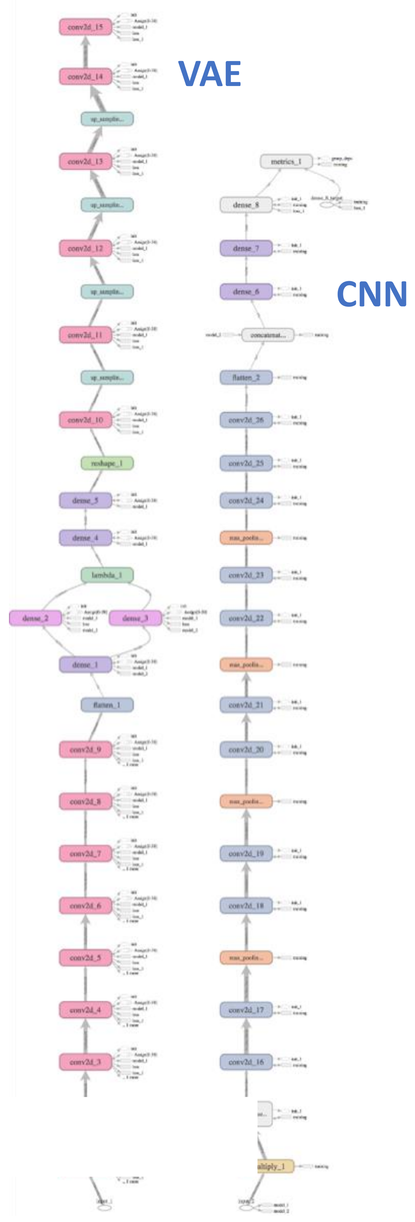
A. VAE1 training minimizes the reconstruction error (cross entropy + Kullback–Leibler (KL) divergence) of the input images where the input and the output are the same image. The hidden layer includes a gaussian sampling to estimate the intensity distribution of input images and the best model is selected based on the minimized error.

B. VAE2 is the pre-trained network from A. Multiplication of output of VAE2 and the input creates an enhanced image that goes through a CNN training process described in the Training section above.

Training and Validation Performance

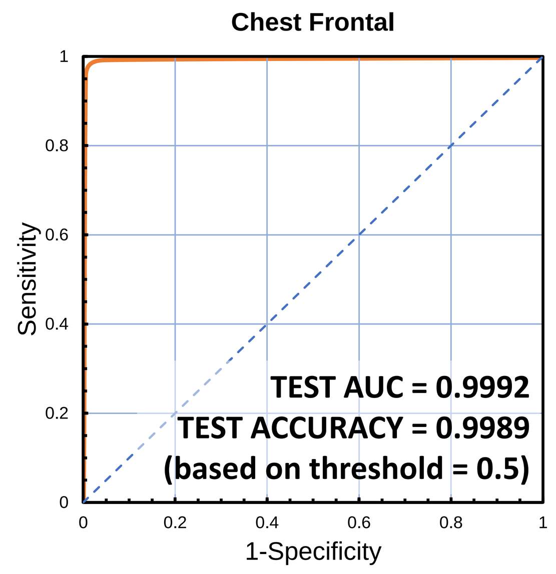


Training loss (upper) and validation loss (cross entropy) (lower), decreased smoothly as accuracy increased, indicating the convergence of the model. Early stopping criteria was employed.



Chest Frontal Detection Model

Test Performance



Confusion Matrix

		AI Predicted View	
		Frontal	Non-Frontal
Actual View	Frontal	2697 (50.2%)	20 (0.4%)
	Non-Frontal	10 (0.2%)	2646 (49.3%)

False Positives

AI Output: Non-Frontal Chest
Ground Truth: Frontal Chest

Images with poor positioning or collimation, as well as uncommon objects in the field of view, were found to be misclassified by AI, likely due to lack of training on such cases.

False Negatives

AI Output: Frontal Chest
Ground Truth: Non-Frontal Chest

Images misclassified by AI included abdominal, extremities with poor collimation, and a mislabeled chest image, likely due to lack of training on such cases.

TASK 2: DETECT CORRECT PATIENT POSITIONING

Data Sources and Number of Images

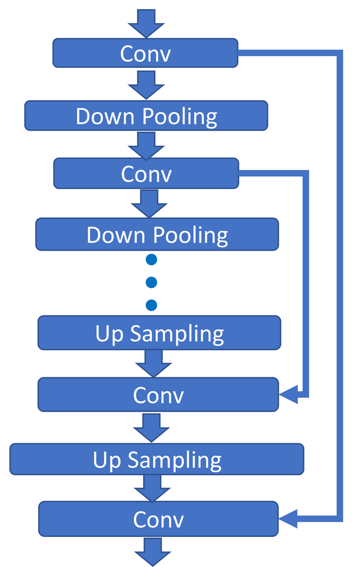
CNN: Canada, USA Hospital 1
Bad Patient Positioning: 1,115; Good Positioning: 1,141

Ratio of Training : Validation : Testing Cases: 70% : 15% : 15%

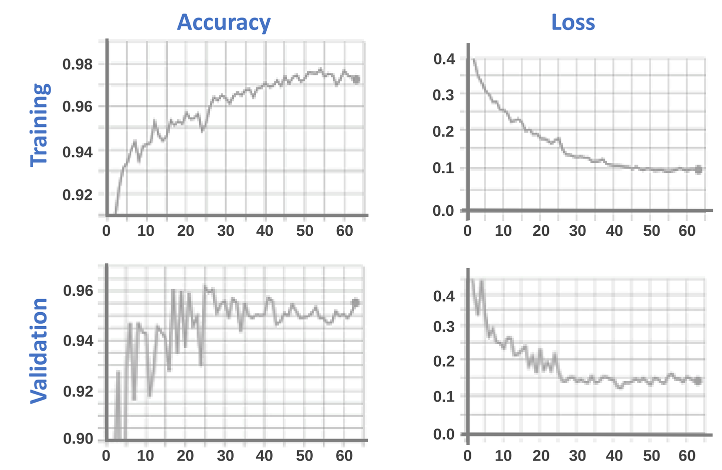
Same Training Protocol as for Task 1:

Network (CNN)

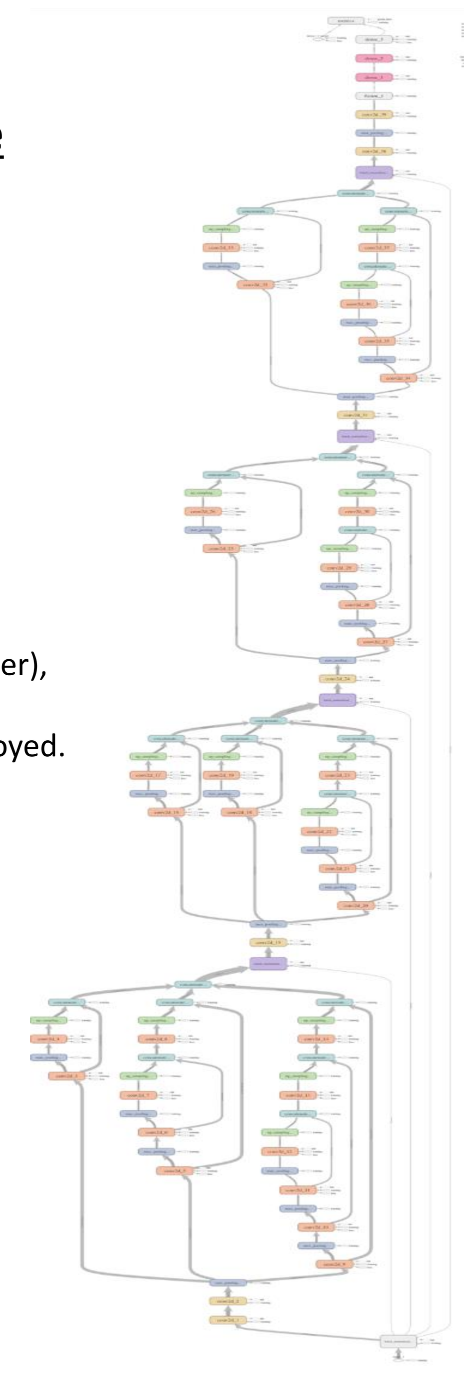
Single Block of our Network:
Each Block is flexible in terms of its depth (number of paired up/down sampling convolutional layers) and the size of convolutional filters (e.g. 3x3, 5x5, 7x7).
Blocks can be connected in series and/or in parallel to form the complete network architecture as shown at right above.



Training and Validation Performance

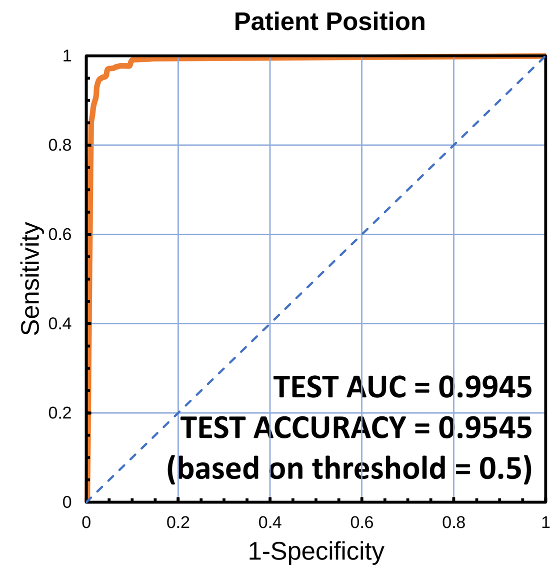


Training loss (upper) and validation loss (cross entropy) (lower), decreased smoothly as accuracy increased, indicating the convergence of the model. Early stopping criteria was employed.



Patient Positioning Detection Model

Test Performance

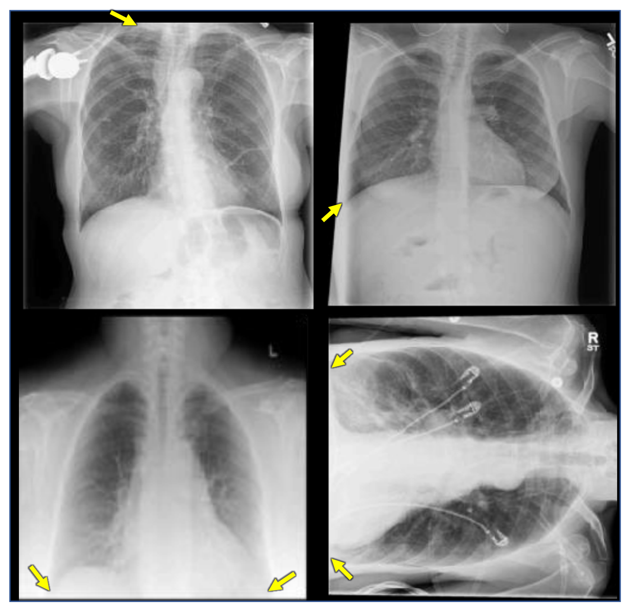


Confusion Matrix

		AI Predicted Positioning	
		Good	Bad
Actual Positioning	Good	160 (48.5%)	7 (2.1%)
	Bad	8 (2.4%)	155 (47.0%)

False Positives

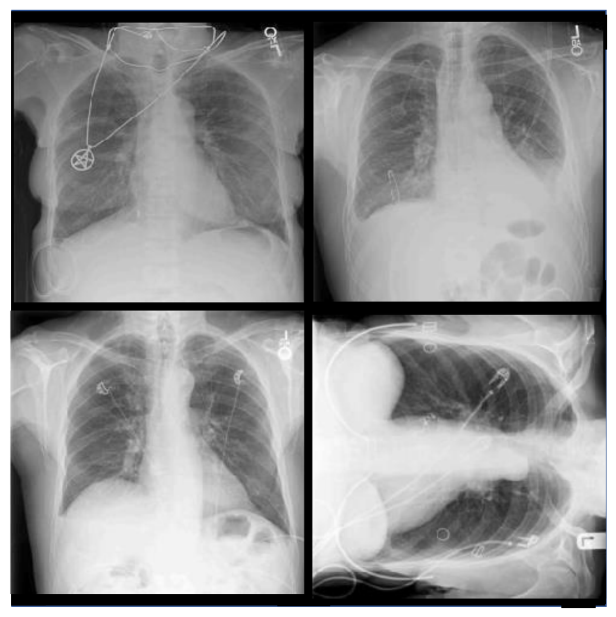
AI Output: Good Positioning
Ground Truth: Bad Positioning



Clipped lung apexes and bases (arrows) were misclassified by AI, more training cases are needed to increase specificity.

False Negatives

AI Output: Bad Positioning
Ground Truth: Good Positioning



Occasional cases were misclassified by AI most likely due to foreign objects in the field of view; reflecting the need for more representative training cases.

CONCLUSION

This work demonstrates the feasibility of using AI algorithms to determine if incorrect anatomy or view were acquired for a AP/PA, as well as whether acceptable patient positioning was achieved for chest x-rays. Further development is warranted to expand anatomy and view types, as well as additional image reject reasons. Algorithms using the same methodology are under evaluation for lateral chest x-rays.
Such AI algorithms would be valuable in a QC program to act as a virtual QC technologist ensuring images are correctly labeled prior to being pushed to the PACS and for automating and improving the image repeat and reject process.

REFERENCES

1. Jones, A. Kyle, Philip Heintz, William Geiser, Lee Goldman, Khachig Jerjian, Melissa Martin, Donald Peck, Douglas Pfeiffer, Nicole Ranger, and John Yorkston. "*Ongoing quality control in digital radiography: report of AAPM Imaging Physics Committee Task Group 151.*" **Medical Physics 42**, 11 (2015): 6658-6670.
2. Pu, Yunchen, Zhe Gan, Ricardo Henao, Xin Yuan, Chunyuan Li, Andrew Stevens, and Lawrence Carin. "*Variational autoencoder for deep learning of images, labels and captions.*" In **Advances in Neural Information Processing Systems**, pp. 2352-2360. 2016.
3. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "*Imagenet classification with deep convolutional neural networks.*" In **Advances in neural information processing systems**, pp. 1097-1105. 2012.