Distributed Deep Learning for Multi-Label Chest Radiography Classification

Maram Monshi, Josiah Poon & Vera Chung

Problem

» **Chest radiography (CXR) supports the diagnosis and treatment for a series of thoracic diseases like pneumonia**

» **Recent automatic classifiers use deep learning but:**
- Neglect label co-occurrence & interdependency
- Fail to make full use of accelerators
- Result in inefficient & computationally expensive models

» **CXR classifiers' performance can be improved by:**
- Leveraging label co-occurrence (Chen et al., 2020)
- Selecting the optimal CXR format (Sabottke and Spieler, 2020)
- Training with an efficient approach

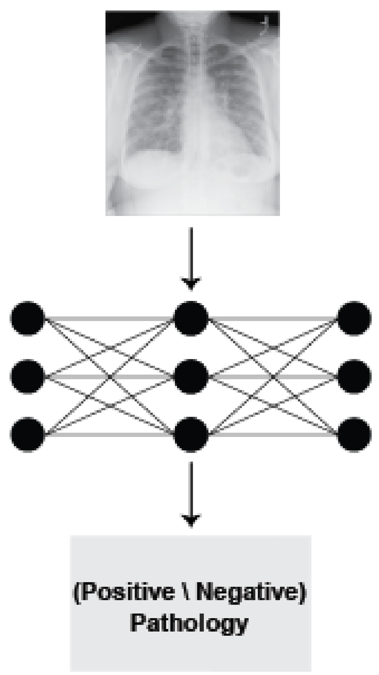


Fig 1. CXR Classification

Literature

» **Binary classification with Convolution Neural Network (CNN)**
- CheXNet (Rajpurkar et al., 2017)
- TieNet (Wang et al., 2018)
- MultiViewModel (Monshi et al., 2019)
- VGG16-based model (Yarnall, 2020)

They did not consider pathology correlation & ignored labels relation

» **Multi-label classification**
- CheXclusion (Seyyed-Kalantari et al., 2020)
- Latent-space self-ensemble (Hou et al., 2021)
- VSEGCN (Hou et al., 2021)

We extended this wave of research using more efficient training methods

Method

» **Dataset:**

1. **MIMIC-CXR** (Johnson et al., 2019)

   377,110 $\xrightarrow{\text{preprocessing}}$ 356,225 CXRs

2. **CheXpert** (Irvin et al., 2019)

   224,316 $\xrightarrow{\text{preprocessing}}$ 212,498 CXRs

- U-zeros method
- Split (80-10-10)

Table 1: Positive label co-occurrence

Label	% of all data	% of label co-occurrence: At	Ca	Co	Ed	EC	Fr	LL	LO	NF	PE	PO	Pa	Px	SD
Atelectasis (At)	16	100	12	6	27	5	4	3	43	0	49	1	2	9	60
Cardiomegaly (Ca)	13	14	100	5	43	7	3	2	48	0	44	1	2	3	58
Consolidation (Co)	7	14	10	100	21	4	3	5	38	0	50	2	7	5	52
Edema (Ed)	25	17	22	6	100	4	2	2	53	0	51	1	2	3	64
Enlarged Cardiom. (EC)	14	18	6	20	20	100	6	5	48	0	36	2	1	7	52
Fracture (Fr)	4	14	9	4	11	7	100	4	40	0	27	3	2	12	40
Lung Lesion (LL)	4	11	7	8	9	6	4	100	58	0	36	3	5	9	35
Lung Opacity (LO)	50	13	12	5	26	5	3	5	100	0	49	2	4	9	58
No Finnding (NF)	11	0	0	0	0	0	0	0	0	100	0	0	0	0	39
Pleural Effusion (PE)	41	19	14	9	31	5	3	4	61	0	100	1	2	8	61
Pleural Other (PO)	2	11	9	9	9	5	8	9	53	0	26	100	4	7	39
Pneumonia (Pa)	3	10	8	17	20	3	2	8	67	0	29	2	100	2	29
Pneumothorax (Px)	9	16	4	4	8	4	5	4	47	0	34	1	1	100	60
Support Devices (SD)	55	17	13	7	29	5	3	3	53	8	46	1	2	10	100

» **Data Augmentation:**

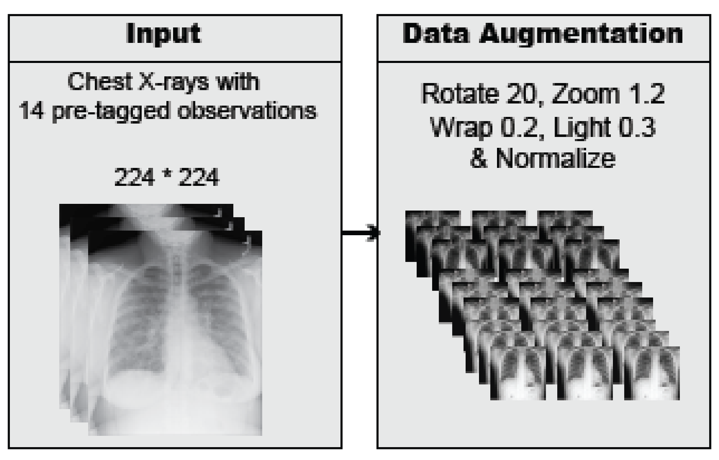
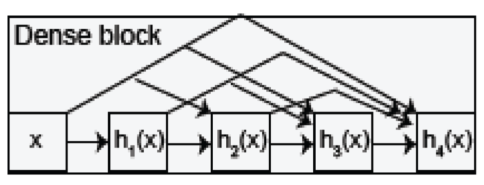


Fig 2. Data Augmentation

To increase the accuracy of detecting abnormalities from CXRs (Monshi et al., 2021)

» **CNN Architecture:**
- Xclassifier is based on Dense CNN (DenseNet-121) (Huang et al.,2017)



$$X_l = H_l([x_0, x_1, \ldots, x_{l-1}])$$

Fig 3. Dense block

Due to its success in CXRs classification (Rajpurkar et al., 2017) (Yao et al., 2017) (Mo and Cai, 2019) (Chen et al., 2020) (Bressem et al., 2020)

» **Antialiasing & Subsampling:**
- Insert a blur kernel $m \times m$ before each downsampling step in DenseNet

$$Relu \circ Conv_{k,s} \rightarrow BlurPool_{m,s} \circ Relu \circ Conv_{k,1}$$

This increases ImageNet (Zhang, 2019) & CXR classification accuracy

Table 2: DenseNet-121 variations models and training performance

Model	Description	Accuracy	AUC
DenseNet-121	Single 7x7 convolution layer with no anti-aliasing layer	90.69	81.34
DenseNet-121d	Three 3x3 convolution layers with no anti-aliasing layer	90.73	81.28
DenseNetblur-121d	**Three 3x3 convolution layers with anti-aliasing blur pool**	**90.80**	**81.96**

» **Distributed Data Parallel (DDP):** (Li et al., 2019)
- Use 64 CXRs (batch size) for each of the 4 GPUs to accelerate convergence

DDP provides 4x speed-up over one GPU & 1.14× to 3.35× speed-up over data parallel

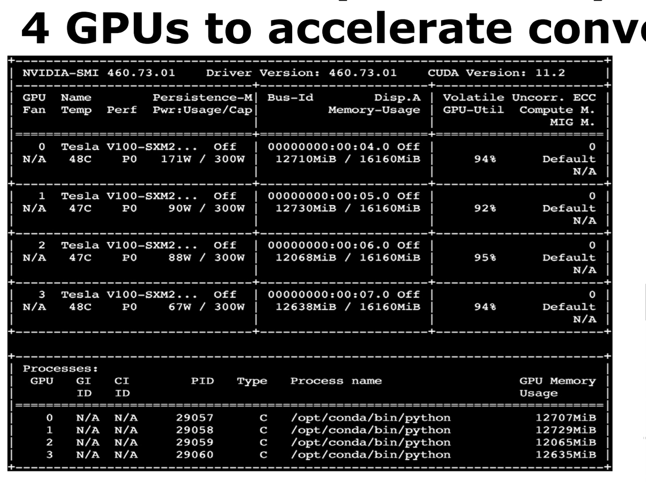
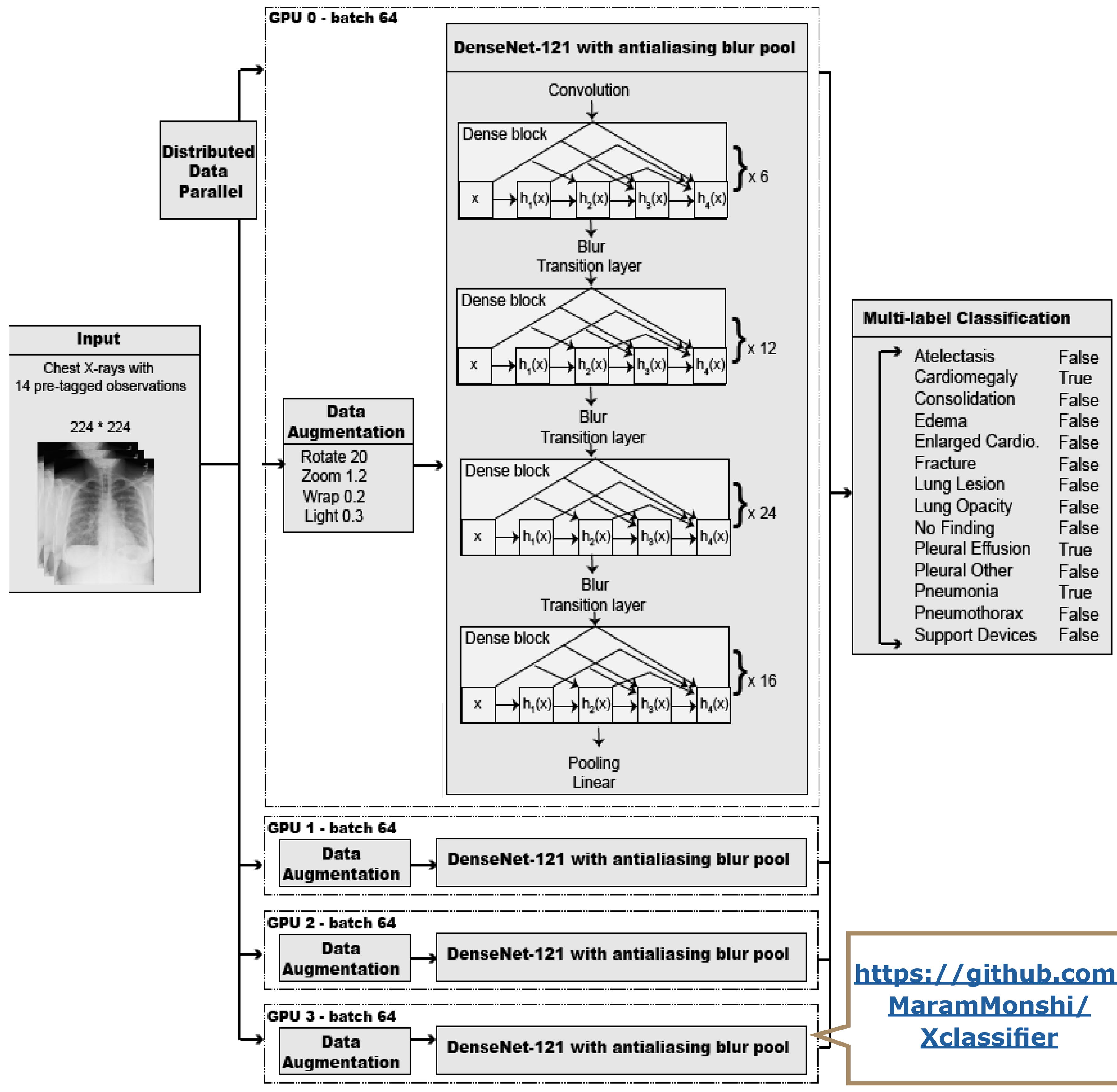


Fig 4. Visualizing DDP

Table 3: Training approaches and training performance

Training Approach	Dataset	Accuracy	AUC	Avg. time per epoch (min)
Single GPU (1 x GPU)	CheXpert	88.09	78.55	16
Data parallel (4 x GPUs)	CheXpert	**88.36**	79.25	14
DDP (4 x GPUs)	**CheXpert**	88.33	**80.10**	**4**
Data parallel (4 x GPUs)	MIMIC-CXR	90.27	80.97	181
DDP (4 x GPUs)	**MIMIC-CXR**	**90.31**	**81.76**	**54**

Implementation



https://github.com/MaramMonshi/Xclassifier

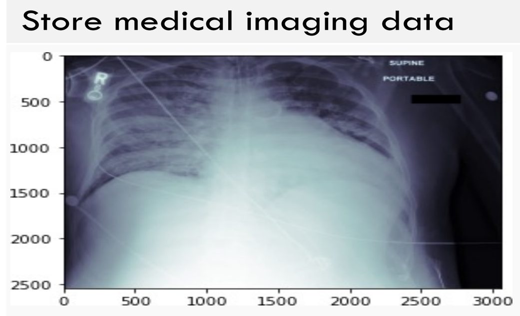
Fig 5. Xclassifier structure

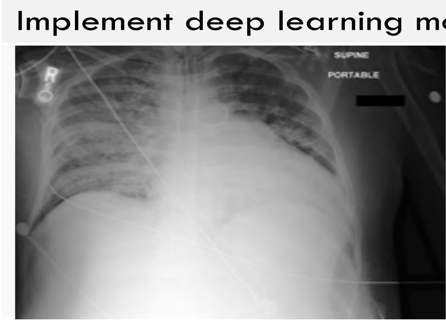
Evaluation

Table 4: Image formats for chest x-rays and training performance

Chest x-ray format	Accuracy	AUC	Avg. time per epoch (min)
DICOM	89.40	80.02	111
JPEG	**89.58**	**81.57**	**6**

Digital Imaging and Communications in Medicine (DICOM)	Joint Photographic Experts Group (JPEG)
Store medical imaging data	Implement deep learning models





DICOM did not improve accuracy & took significantly more time to train than JPEG

Table 5: Comparing the Xclassifier with the benchmark

Multi-label classifier	Dataset	Accuracy	AUC
Latent-space self-ensemble (Gyawali et al., 2019)	CheXpert	_	66.97
CheXclusion (Seyyed-Kalantari et al., 2020)	CheXpert		80.50
Xclassifier	**CheXpert**	**89.61**	**83.89**
VSE-GCN (Hou et al., 2021)	MIMIC-CXR	_	72.10
CheXclusion (Seyyed-Kalantari et al., 2020)	MIMIC-CXR	_	83.40
Xclassifier	**MIMIC-CXR**	**92.17**	**84.10**

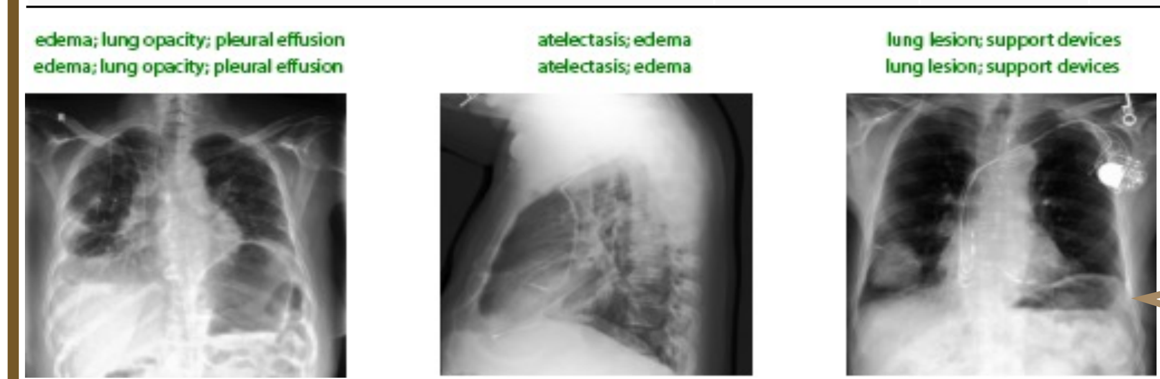


Fig 6. Correct output sample by the Xclassifier Model

Xclassifier improves multi-label classification performance by 0.70% AUC on MIMIC-CXR & by 3.39% AUC on CheXpert

Conclusion

» **Contribution:**
- Propose Xclassifier, an efficient multi-label classifier that trains enhanced DenseNet-121 with blur pooling to detect 14 observations from CXRs
- It accomplishes an ideal memory utilization, GPU computation, & high AUC on two large chest radiography datasets, MIMIC-CXR & CheXpert

» **Future Work:**
- Investigate the use of DICOM in detecting diseases with small & complex structures to offer a greater degree of understanding of our initial findings
- Concatenate patient data like age and gender to the flattened layer to improve prediction

» **References:**

Bressem et al., (2020). Comparing different deep learning architectures for classification of chest radiographs.
Chen et al., (2020). Label co-occurrence learning with graph convolutional networks for multi-label chest x-ray classification.
Gyawali et al., (2019). Semi-supervised learning by disentangling and self-ensembling over stochastic latent space.
Hou et al., (2021). Multi-label learning with visual-semantic embedded knowledge graph for diagnosis of radiology imaging.
Huang et al., (2017). Densely connected convolutional networks.
Irvin et al., (2019). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison.
Johnson et al., (2019a). MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports.
Li et al., (2020). PyTorch Distributed: Experiences on Accelerating Data Parallel Training.
Mo & Cai (2019) Deep learning based multilabel chest x-ray classification with entropy weighting loss.
Monshi et al., (2021). CovidXrayNet: Optimizing Data Augmentation and CNN Hyperparameters for Improved COVID-19 Detection from CXR.
Monshi et al., (2019). Convolutional neural network to detect thorax diseases from multi-view chest x-rays.
Rajpurkar et al., (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning.
Sabottke et al., (2020). The effect of image resolution on deep learning in radiography.
Seyyed-Kalantari et al., (2020). CheXclusion: Fairness gaps in deep chest X-ray classifiers.
Wang et al., (2018). Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays.
Yao et al., (2017). Learning to diagnose from scratch by exploiting dependencies among labels.
Yarnall (2020). X-Ray Classification Using Deep Learning and the MIMIC-CXR Dataset.
Zhang (2019). Making convolutional networks shift invariant again.