Chest X-Ray Disease Diagnosis Model Comparison

Youtube: https://youtu.be/szQPOGJ4dvw

Github: https://github.com/pwodarz/CS598-DL4H-Chest-X-ray-Classification

Pierson Wodarz
UIUC
wodarz2@illinois.edu

Steve Su
UIUC
steven36@illinois.edu

Atashi Rana
UIUC
atashir2@illinois.edu

ABSTRACT

In this paper, we investigate several deep learning models for use in diagnosing diseases from chest X-ray images. We train a simple convolutional neural network (CNN), a ResNet model, a DenseNet model, and an Extended CNN model on a subset of images from the CheXpert dataset of chest X-ray images, both for binary classification of pneumonia and multi-label classification of multiple pathologies. Our Extended CNN model is based on our Simple CNN model but includes features that are not commonly used in X-ray classification, namely multiple image views and electronic health record (EHR) data. Our goal is to see whether these additional features can give researchers another option in which to improve their models. In this study we evaluate the results of our Extended CNN model using accuracy, AUC, and F1 score and compare to our Simple CNN model as the baseline and also to the state-of-the-art models developed by other researchers. Both the ResNet and DenseNet were modified for our needs and did not use pre-trained weights. We find that we are able to achieve and replicate similar performance using these state-of-the-art models with our dataset. The results from our Extended CNN model fell short of those of the state of the art models, however with some modifications we believe there could be some merit with the framework.

Keywords

Deep learning, Healthcare, ResNet, DenseNet, X-ray

1. INTRODUCTION

X-ray images have a wide array of applications, from diagnosing diseases and fractures to surgical planning and instrument guidance during operation [1]. In the UK alone (population 68 million) 23.2 million X-rays were reported to have been performed in the 2019/2020 reporting year [2]. The large number of X-rays performed presents an opportunity for major cost savings using deep learning techniques to identify and diagnose diseases. Additionally, these diagnoses may be performed faster and more accurately. Increasing the accuracy and speed of diagnoses will lead to better treatment and patient outcomes which, in turn, further reduces costs.

Several researchers have created deep CNN networks [3], image segmentation [4], and CNN networks which include spatial information of the disease artifact [5][6] for use in X-ray disease diagnosis and image classification, achieving state-of-the-art results which begin to approach the performance of trained radiologists. Rajpurkar et al. [3] used a 121 layer CNN model to classify 14 different pathologies using frontal chest X-rays. The model had a margin of > 0.05 AUROC over previous state of the art results and had statistically significant higher performance than average radiologists.

Following in these researchers' footsteps, we utilized and trained eight total models: A simple CNN, an Extended CNN, a ResNet model, and a Densenet model, which we first programmed for binary classification of pneumonia and then extended to multi-label classification of 14 common chest radiographic observations. We trained these models on a subset of data from the CheXpert dataset. The full dataset contains 224,316 images labeled for 14 common chest radiographic observations from 65,240 patients.

We find that we are able to meet the performance of state-of-the-art models developed by other researchers, which implies a performance similar-to or exceeding trained radiologists. This result holds for both binary classification and multi-label classification, though for multilabel this is restricted to only certain pathologies.

2. APPROACH

2.1 Data

Of foremost importance to creating an image classification model is collecting labeled data for use in training the model. In this paper, we utilize the CheXpert dataset [7]. The full dataset contains 224,316 images labeled for 14 common chest radiographic observations from 65,240 patients.These labels were generated using a labeling program trained on radiologist-labeled reports which then extracted observations from free-text radiology reports to generate the labeling for the full dataset.

For each datapoint, the dataset contains an image, sex, age, view perspective, and labeled list of 14 observations (labeled 0 for negative, -1 for uncertain, 1 for positive, and blank for unmentioned). Note that the labels are multi-label, meaning that each image can have multiple labels for the different pathologies (e.g. be both an edema and a consolidation).

Label	Pathology Name
0	No Finding
1	Enlarged Cardio mediastinum
2	Cardiomegaly
3	Lung Opacity
4	Lung Lesion
5	Edema
6	Consolidation
7	Pneumonia
8	Atelectasis
9	Pneumothorax
10	Pleural Effusion
11	Pleural Other
12	Fracture
13	Support Devices

Table 1: Mapping of multi-labels and the corresponding pathology names

For the purposes of processing the data and training the model, we utilized a subset of the full dataset due to limitations in storage and computation. In particular, for pneumonia classification, we select a subset of 1,055 patients, which includes 1,268 images explicitly labeled (positive or negative) for pneumonia. For multilabel classification, we select a subset of 850 patients, which included 2,787 images. These datasets were arrived at by progressively increasing the amount of data until our resource limitations were hit. In some cases the training data set for the Extended CNN model had fewer positive labels because some samples did not have both frontal and lateral images.
To standardize the images for feeding into our model, we centercrop the images and downscale to 224x224 as the image sizes are not consistent across all images. We also find that the smaller size aids with processing.
We utilize the same dataset for training and testing as well as the same transformations and data pre-processing for all of our models to obtain a comparable level of performance between them.

2.2 Metrics

Commonly used metrics in existing literature includes F1 and AUC. Some researchers utilize AUC [8] as their primary performance metric, while others have utilized F1 score [9] to characterize performance, and others use F1, AUC, and accuracy [10]. To provide comparability between the models in this paper, as well as those presented by other researchers, we will output the accuracy, AUC, and F1 of each of the models.
Accuracy will provide a more approachable metric that can serve to orient the reader to the general performance. F1 serves to present the performance in terms of both precision and recall, each of which is of high importance in healthcare when diagnosing disease. AUC also balances the positive and negative classes. All three allow for comparison between different models, which we do in this paper.
The standardized dataset used in this paper will allow for a more accurate comparison between the models described in this paper using the aforementioned metrics. However, the metrics can still provide a loose guide for the comparison of performance of models described in this paper to those generated by other researchers.

2.3 Models

For the first part of our investigation we create binary classification neural network models: a simple CNN 6 layer model, an 18-layer ResNet model, and a 121-layer DenseNet model, and an Extended CNN model. The simple CNN will serve as the baseline model and the ResNet and DenseNet will serve as our state of the art models to compare with our Extended model. To extend each of these binary classification models to handle multi label classification, we add a final fully connected layer with 14 neuron outputs on top of each model. We apply a sigmoid nonlinearity function to output the probability for each label. This required trivial changes to the code in the form of the dataloader and the model architecture. The pytorch library v1.8.1 is used to construct these models.

2.3.1 Simple CNN

The simple CNN is built from scratch to simulate a naive approach to the problem and serve as a baseline for performance comparisons. The model consists of three 2d convolution layers with outputs of 8, 16 and 32 feature maps. Each 2d convolution layer was followed by a max pool layer and a leaky relu activation function. To improve learning we added batch normalization between the layers. Next the output of the convolution network was flattened into a vector and passed to the classification network which consisted of three fully connected layers. Each fully connected layer was followed by a dropout layer and an activation function. A factor of 50% percent was chosen in each drop out layer. The aim of the dropout layer is to prevent overfitting during training by randomly dropping connections between the layers [11]. A leaky relu function was used as the activation function for the first two fully connected layers and a sigmoid for the last to give the final probability assignment of the binary classification. The loss function was binary cross entropy and the optimization function was Adam with a learning rate of 0.001.

2.3.2 ResNet Model

The ResNet model has been used for X-ray image classification in previous studies [12] and has been used recently to classify COVID-19 in chest X-rays [13], supporting its use in image classification, and with X-ray data in particular. ResNet seeks to improve training for very deep neural networks, which have historically been difficult to train but provided leading

performance on existing image classification tasks [14, 15]. ResNet uses residual learning to ease optimization and allows for increased depth which leads to improved accuracy [16]. The key component of the ResNet architecture is skipped connection, which skips training and connects directly to the output, such that the network first the residual mapping. In our case, we used ResNet 18 which is an 18 layer plain network architecture with the skip connections added. Similar to the simple CNN, a binary cross entropy loss function was utilized, with a Stochastic Gradient Descent optimizer at a learning rate of 0.001.

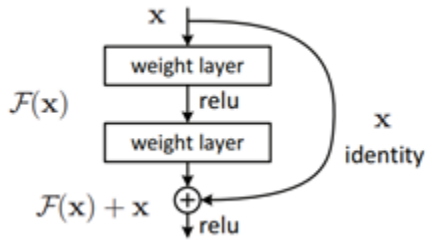

Image 1: ResNet skip connection architecture

2.3.3 DenseNet Model

The DenseNet model was one of the approaches for use in X-ray image classification pioneered by Rajpurkar et al. which achieved performance beyond that of the average radiologist [3]. This model has been used effectively for the multi-label classification of chest X-ray images [17]. Huan et al. [18] describe the DenseNet architecture that consists of 3 key components that are the essential parts of the algorithm: input and output layers, transition Layer, and dense blocks. All layers are interconnected in the DenseNet model. At an individual layer, the feature maps of all preceding layers are considered as inputs which imply that the feature maps of each layer are concatenated. The transition layer is used for driving scalability, and the dense blocks connect all layers. Each layer gets its input from all the preceding layer's output. The DenseNet model helps to build deeper neural networks. It also makes it efficient to train by using the shorter connections between the layers.

In our case, we use DenseNet 121 which is composed of a 121 layer architecture. To train, we use a learning rate of 0.001 with an Adam optimizer and use a binary cross-entropy loss function similar to the other models.

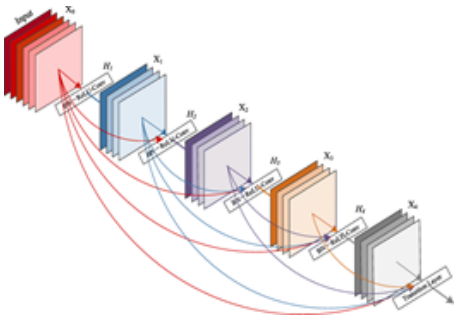

Image 2: DenseNet architecture depiction

2.3.4 Extended CNN Model

Our final model, Extended CNN, attempts to improve the simple CNN architecture by incorporating lateral chest views and EHR data. One limitation in our study is the EHR data is quite limited, only age and sex are available. However, future data could be introduced into the model after the framework has been established.

To create this hybrid model, we will now have two images as input, frontal and lateral views. To make sure the image selection is applicable, the images have to be filtered from the repository so that the two images come from the same patient and the images were taken at the same time. Since each image has truth labels assigned to it, for each image pair we will merge the data. The new hybrid model now will have two input channels instead of one. The rest of the convolution network will remain the same as described in the Models section.

To include the EHR data we concatenate this data in the classification portion of the network. This occurs after the third convolution layer where the output is flattened to a vector before being passed to the classifier. The EHR data gets concatenated to this vector at this point.

Before the EHR data can be concatenated it needs to be normalized. For male or female this is trivial. For age we normalize to continuous values between 0 and 1. The rest of the classifier is the same with the exception of expanding the input for the first fully connected layer for these two additional features. For the multi-label model, similar changes were made as discussed in the Simple CNN model.

2.4 Model Training/Validation

The cloud platform used to run our models was Google Colab. This allowed us to train our models on a GPU.

None of the models were pre-trained, and were trained only on the CheXpert training dataset. The same training and testing image sets were used for training each model, as well as image transformations and preprocessing.

For binary classification, we classify the images by the pneumonia diagnosis. To do so, we restrict the training and testing set to images with an explicit positive or negative pneumonia indicator. This results in a total of 1,146 images for training and 122 images for testing. There were no overlapping patients between these training and testing sets. This is to avoid testing on an image for a patient when another image from that same patient may have been used to train the models.

We train these models until the loss plateaus or we hit practical limitations with resources (training time was limited per model). Extending the models to multi-label requires retraining with the same methods as described above, but with all labels and data from the training dataset input. For unlabeled pathologies, we made the assumption that the patient did not have the associated diagnosis and classified it as an explicit negative. Additionally, to train on uncertain input, we collapsed the classification for all

uncertain labels to be negative. The multilabel models were trained with 2,473 images and tested on a set of 314 images.

Model	Epochs	Final Loss	Training Time
Binary Extended CNN	50	0.61	1 hour
Binary Simple CNN	50	0.598	1 hour
Binary ResNet18	100	0.035	6 hours
Binary DenseNet121	40	0.093	9 hours
Multilabel Extended CNN	50	0.321	1 hour
Multilabel Simple CNN	50	0.339	1 hour
Multilabel ResNet18	40	0.321	3 hours
Multilabel DenseNet121	40	0.339	7 hours

Table 2: Model training metrics

3. RESULTS

3.1 Binary Classification

Model	Accuracy	AUC	F1
Extended CNN	0.347	0.591	0
Simple CNN	0.639	0.61	0.778
ResNet18	1	1	1
DenseNet121	0.791	0.954	0.868

Table 3: Accuracy, AUC, and F1 scores for binary classification of X-ray images for the pneumonia pathology

The Extended CNN model has a very low accuracy score compared to the rest of the models. It only achieves a score of 0.347 which is well below our baseline Simple CNN model, whose score is 0.639. However, the Extended CNN's AUC score at least approached the value of the Simple CNN, 0.591 versus 0.61 respectively. Lastly, the Extended CNN's F1 score is zero. This indicates that it did not predict any true labels in the validation set. Not surprisingly, the other two established models do much better.

Of particular note are our scores for ResNet18 which, after training for 100 epochs, is able to achieve perfect scores for accuracy, AUC, and F1. To validate this result further, we applied the model to 631 images which were unlabeled for pneumonia, an assumed negative, and examined the results. We found that we maintained 1.0 for all scores. Given that the same test and train sets were used for all three binary models, we can be assured that this strong performance is not due to errors in our code. This lends strong credence to the results, having been tested on over 753 images. This performance is to be considered state-of-the-art given that there is no room for improvement.

Our DenseNet model achieved mediocre performance, particularly in terms of accuracy and F1. However, it still exceeded the performance of a simple CNN and had results comparable to historical research. DenseNet scores improved over a naive approach, but did not approach an overall level of performance expected. While there may be utility to applying this model on this image classification task, other models may perform better and therefore be more useful in practice.

The lower F1 score of DenseNet compared to ResNet is also replicated in other studies [20], lending support to the use of ResNet for these types of image classification tasks.

3.2 Multi Label

Please refer to Table 1 for the mapping of each label and its corresponding pathology name.

In some cases the F1 score is zero, most notably for the extended and Simple CNN models. In these cases our model did not predict any cases of the positive label. Because of this we primarily use accuracy and/or AUC as our performance metric. To help make sense of the data both accuracy and AUC was

	Extended CNN			Simple CNN			ResNet			DenseNet		
Label	**Acc**	**AUC**	**F1**	**Acc**	**AUC**	**F1**	**Acc**	**AUC**	**F1**	**Acc**	**AUC**	**F1**
0	0.81	0.684	0	0.904	0.785	0	0.897	0.819	0.015	0.897	0.749	0.008
1	0.946	0.644	0	0.929	0.521	0	0.956	0.666	0	0.956	0.587	0
2	0.895	0.733	0	0.843	0.62	0	0.875	0.757	0.006	0.875	0.656	0
3	0.673	0.571	0	0.64	0.687	0.638	0.614	0.669	0.623	0.584	0.643	0.372
4	0.914	0.437	0	0.958	0.607	0	0.96	0.714	0	0.96	0.586	0
5	0.937	0.623	0	0.748	0.722	0	0.775	0.754	0.283	0.769	0.714	0.144
6	0.971	0.856	0	0.92	0.574	0	0.938	0.654	0	0.938	0.62	0
7	0.965	0.478	0	0.971	0.543	0	0.977	0.622	0	0.977	0.644	0
8	0.917	0.669	0	0.872	0.579	0	0.857	0.634	0	0.858	0.641	0.006
9	0.946	0.664	0	0.942	0.561	0	0.893	0.737	0.282	0.895	0.674	0
10	0.743	0.743	0.308	0.643	0.665	0.504	0.672	0.739	0.615	0.64	0.682	0.336
11	0.959	0.485	0	0.984	0.576	0	0.984	0.754	0	0.982	0.746	0.435
12	0.949	0.527	0	0.933	0.657	0	0.96	0.696	0	0.96	0.629	0
13	0.695	0.581	0	0.646	0.68	0.725	0.633	0.744	0.73	0.649	0.685	0.699
Average	**0.88**	**0.621**	**0.022**	**0.852**	**0.627**	**0.133**	**0.857**	**0.711**	**0.182**	**0.853**	**0.661**	**0.143**

Table 4 Multilabel Accuracy, AUC, and F1 metrics for Extended CNN, Simple CNN, ResNet, and DenseNet

averaged across all 14 labels for each model (Figure 1). On observation it appears that our Extended Model performed well with the highest accuracy but when looking at AUC score it performed the worst. Since our model tends to be biased predicting negative labels (as evidenced by many F1 scores to be zero) and unbalanced validation set, the accuracy score can be misleading. More details are included in the discussion section. For this reason, AUC is a better metric for this case. When we take this into account the ResNet has the best performance with an AUC score of 0.711. Following are DenseNet, 0.611; Simple CNN, 0.627; and Extended CNN, 0.621.

It is interesting to note that in some isolated cases the Extended CNN performs better than the ResNet, particularly for the Consolidation label #6 (Figure 2). The AUC score was 0.856 for the Extended model and 0.654 for the ResNet model. More rigorous testing is needed to confirm validity of this.

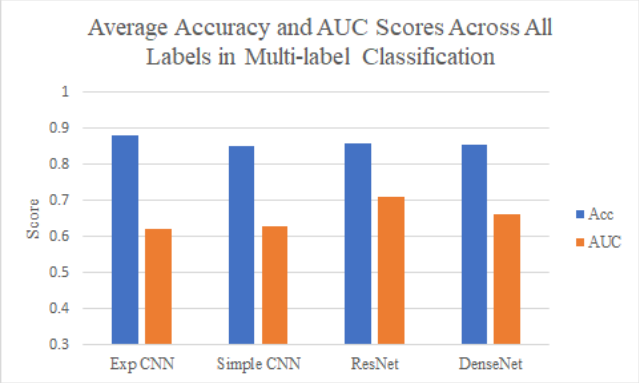

Figure 1. Average Accuracy and AUC scores for each model

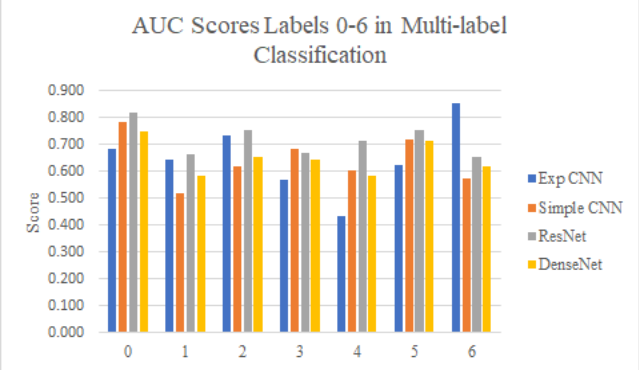

Figure 2. AUC scores for each model and labels 0-6

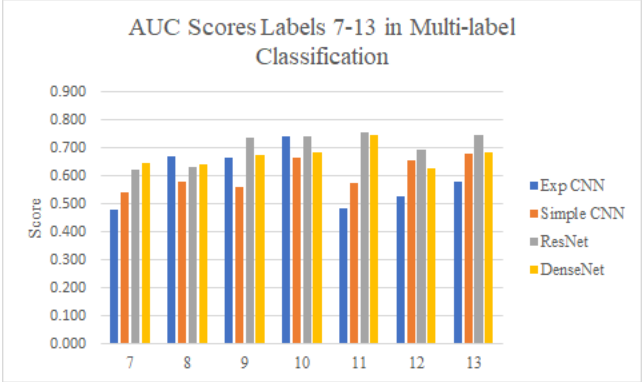

Figure 3. AUC scores for each model and labels 7-13

4. DISCUSSION

With respect to the binary classification of pneumonia, the results from the Extended CNN model do not appear to be any better than the Simple CNN and falls short of the ResNet and DenseNet by a considerable amount. The accuracy is particularly poor at 0.347, whereas the baseline Simple CNN's score is 0.639. If the AUC is taken into account the two model's performance seems closer. The scores for the Extended Model and Simple CNN are 0.591 vs 0.610, respectively (Table 3). When reviewing the F1 score, the Extended CNN has a score of zero which indicates the model predicts all of the labels to be false suggesting a higher bias towards false predictions versus the Simple CNN model. While the accuracy looks very poor, the AUC is a more representative score, especially when we look at the multi-label results.

With respect to the average multi-label accuracy scores (average across all 14 categories) (Figure. 1) it would seem the Extended Model performed the best having the highest score. However, the Extended model seems to be biased in mostly predicting a false label, indicated by F1 scores of zero. This factor, combined with an unbalanced validation dataset (more false labels than true) contributes to misleading accuracy scores.

To help make sense of each model's overall performance an average AUC score of the 14 labels is used (Figure 1). In this case, ResNet has the highest score followed by DenseNet and then Simple CNN and Extended CNN. This makes more sense as we know both ResNet and DenseNet are state of the models. When looking at the Extended CNN and the Simple CNN, their average AUC scores are almost the same, 0.621 versus 0.627, indicating that the two models perform similarly.

One of the possible reasons why the Extended CNN model did not show overall improvements could be fewer positive training samples. In some categories, namely pneumonia, the training set used for the Simple CNN, ResNet, and DenseNet did not always have an accompanying lateral view. In these cases that observation was removed since the Extended model requires both lateral and frontal views. Recall, the other models only required a

frontal view. Secondly, the EHR data that we had was quite limited. Only sex and age were available to us. Nonetheless we decided to include it and provide a framework which can be expanded with the availability of additional data.

Even with these setbacks we feel the primary reason for poor performance in our Extended CNN model is due to poor learning. We observed that the loss plateaued fairly quickly, typically within the first 10 epochs. Thereafter only very minute improvements were seen. This is also a factor in the Simple CNN model. Some attempts were made to optimize the performance through learning rate and number of output feature maps but performance was little changed. Turning to other possible causes, deep networks suffer the problem of exploding and vanishing gradients which prevents learning where the weights of the network no longer get updated. During training, neural network's weights are updated with each backpropagation iteration. In deep networks the gradient can become extremely small or large, once this happens subsequent updates to the weights get pegged (to a very high or low number) and ceases to effectively update its value, stagnating the learning. Our simple network is relatively small but we surmised that our network could be affected to some degree. To overcome this problem researchers have proposed a normalization process and leaky Relu activation functions to help reduce stagnating learning [19]. Based on this we implemented a 2d batch normalization after each CNN layer and a 1d batch normalization after each feed forward layer in the classification network. Additionally we used leaky Relu activation functions instead of normal Relu functions, however there were only marginal gains.

Unlike the Extended and Simple CNN the ResNet and DenseNet did not have stagnating loss values. Even after the maximum time we allotted for training these models still showed decreasing progression of loss values. To ensure good learning performance these state of the art models have unique features which reduce the likelihood of exploding/vanishing gradients while still having a deep network for good feature detection. To do this, the ResNet model uses skip connections in an effort to preserve information. In the case of DenseNet it does something similar to ResNet in that it passes information by skipping layers. One difference is that during forward propagation, the output of each CNN layer is not only passed to its immediate neighbor but to every CNN layer downstream. As a result, every output of a layer becomes the input for all subsequent layers. With this architecture information can be preserved during end to end training, alleviating vanishing/exploding gradients [18]. Perhaps with these ideas our Extended CNN model could be improved beyond what the batch normalization and leaky relus were able to do for our model.

Even though the Extended CNN model did not perform well overall, it did show possible performance benefit over the other models in isolated cases. As mentioned in the results section its AUC score was the highest for label #6 Consolidation. It may be possible that having the lateral view and/or EHR data really makes a performance difference for certain labels. More studies are needed to validate this.

5. CONCLUSION

Although the features introduced in our Extended Model did not improve overall performance, there may be isolated cases where it performs better. Some of the AUC scores in the Extended CNN model were on par with the Resnet model for the multi-label classification task. It would seem reasonable that adding additional views and EHR data would be of great benefit. For example there may be cases where the disease artifact can only be seen in one view. Regarding EHR data, symptoms can help decipher the disease a patient has. Physicians seldom make decisions using only one source of information especially in difficult corner cases.

In order to make the Extended CNN model a useful overall diagnostic tool we need to improve the learning. By borrowing some of the concepts of the ResNet and DenseNet models perhaps this can be achieved. Also, more EHR data is necessary to be effective in supporting the classifier. In this study we were limited to just sex and age. Physician notes and medical codes would be a great source of additional data.

If the problem of poor learning can be overcome it is possible that the Extended CNN could be a useful tool in chest X-ray disease classification, especially in difficult corner cases by the inclusion of a more diverse set of data.

6. REFERENCES

[1] Center for Devices and Radiological Health. (n.d.). Medical x-ray Imaging: FDA. Retrieved March 24, 2021, from https://www.fda.gov/radiation-emitting-products/medical-imaging/medical-x-ray-imaging

[2] *Diagnostic Imaging Dataset Annual Statistical Release 2019/20* (Rep.). (n.d.). Retrieved March 23, 2021, from NHS England website: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2020/10/Annual-Statistical-Release-2019-20-PDF-1.4MB.pdf

[3] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225, 2017.

[4] Hesamian MH, Jia W, He X, Kennedy P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. J Digit Imaging. 2019 Aug;32(4):582-596. doi: 10.1007/s10278-019-00227-x. PMID: 31144149; PMCID: PMC6646484.

[5] Gündel, Sebastian & Grbic, Sasa & Georgescu, Bogdan & Liu, Siqi & Maier, Andreas & Comaniciu, Dorin. (2019). Learning to Recognize Abnormalities in Chest X-Rays with Location-Aware Dense Networks: 23rd Iberoamerican Congress, CIARP 2018, Madrid, Spain, November 19-22, 2018, Proceedings. 10.1007/978-3-030-13469-3_88.

[6] Z. Li, C. Wang, M. Han, Y. Xue, W. Wei, L.-J. Li, and L. Fei-Fei. Thoracic disease identification and localization with limited supervision. CVPR, 2018.

[7] Irvin, Jeremy, et al. "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. No. 01. 2019.

[8] Bassi, Pedro RAS, and Romis Attux. "COVID-19 detection using chest X-rays: is lung segmentation important for generalization?." arXiv preprint arXiv:2104.06176 (2021).

[9] Öksüz, Coşku, Oğuzhan Urhan, and Mehmet Kemal Güllü. "Ensemble-CVDNet: A Deep Learning based End-to-End Classification Framework for COVID-19 Detection using Ensembles of Networks." arXiv preprint arXiv:2012.09132 (2020).

[10] Han, Yan, et al. "Pneumonia Detection on Chest X-ray using Radiomic Features and Contrastive Learning." arXiv preprint arXiv:2101.04269 (2021).

[11] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov; 15(56):1929−1958, 2014."Dropout: A Simple Way to Prevent Neural Networks from Overfitting". Jmlr.org. Retrieved July 26, 2015.

[12] Vianna, Vinicius Pavanelli. "Study and development of a Computer-Aided Diagnosis system for classification of chest x-ray images using convolutional neural networks pre-trained for ImageNet and data augmentation." arXiv preprint arXiv:1806.00839 (2018).

[13] Nwuso, Lucy, et al. "Semi-supervised Learning for COVID-19 Image Classification via ResNet." arXiv preprint arXiv:2103.06140 (2021).

[14] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[15] R. K. Srivastava, K. Greff, and J. Schmidhuber. Training very deep networks. 1507.06228, 2015.

[16] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[17] Jiashi Zhao1 , Mengmeng Li1 , Weili Shi1 , Yu Miao1*, Zhengang Jiang1*, Bai Ji2 A deep learning method for classification of chest X-ray

[18] Huang, G., Liu, Z., Van der Maaten, L. And Weinberger, K. Q. Densely Connected Convolutional Networks. https://research.fb.com/wp-content/uploads/2017/05/densely-connected.pdf

[19] Zhenwei Dai, Reinhard Heckel: Channel Normalization in Convolutional Neural Network avoids Vanishing Gradients. CoRR abs/1907.09539 (2019)

[20] Mamalakis, Michail, et al. "DenResCov-19: A deep transfer learning network for robust automatic classification of COVID-19, pneumonia, and tuberculosis from X-rays." arXiv preprint arXiv:2104.04006 (2021).

7. CONTRIBUTIONS

Pierson Wodarz
- Initial dataloader and setup
- Binary ResNet
- Binary DenseNet
- Presentation
- Associated paper sections

Steve Su
- Binary Simple CNN
- Binary Extended CNN
- Overall Multilabel extension of models
- Multilabel Simple CNN
- Multilabel Extended CNN
- Challenges and and learnings
- Associated paper sections

Atashi Rana
- Multilabel ResNet
- Multilabel DenseNet
- Summary PPT Slides
- Associated paper sections