Proceedings of 13th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2023 Month: April Volume: 13

Chest X-Ray Classification using DenseNet

Ankit Karna [a], Aadarsh Jha [b], Alish Dahal [c], Anup Pandey [d], Tantra Nath Jha [e]

[a, b, c, d, e] *Department of Electronics and Computer Engineering, Purwanchal Campus, IOE, TU, Nepal*

[a] karna.ankit1@gmail.com , [b] aadarshjha87@gmail.com, [c] alishd0225@gmail.com, [d] pandeyanup58@gmail.com, [e] tnjha@ioepc.edu.np,

Abstract

Chest X-Ray classification is a challenging and time consuming task in medical image classification due to the complexity of the human chest structure and the subtle variations in X-Ray images caused by different medical conditions. This paper presents a model, DCXNet that can detect 14 different chest conditions from Chest X-Rays. The model makes use of deep learning techniques and transfer learning methods for better accuracy and faster training time. The presented model, DCXNet is a customized DenseNet-169 model which is a 169 layer Convolutional Neural Network (CNN) trained on ChestX-Ray14, one of the largest publicly available Chest X-Ray dataset, containing over 112,000 frontal view X-Ray images of 30,805 unique patients with 14 chest conditions. This dataset was obtained from National Institutes of Health (NIH), USA. Various tools like NumPy, Pandas were used for initial data analysis, while Matplotlib and Seaborn was used for data visualization. The presented model was implemented in Tensorflow.

The presented model is compared to other existing models on the basis of AUC metric. The mean AUC of the presented model is 0.82 which outperformed Wang (AUC ≥ 0.09) and Yao (AUC ≥ 0.02) and acheived on par performance with CheXNeXt (AUC ≤ 0.02). The training time was much faster than CheXNeXt which took around 20 hours for each training stage while the presented model DCXNet completed training in just 4 hours.

Keywords

X-Ray, CNN, DenseNet-169, Transfer Learning, Tensorflow

1. Introduction

The most used imaging test worldwide is chest radiography, which is essential for the early detection, diagnosis, and treatment of many serious illnesses. Chest illnesses are a major public health concern in the country. Radiology is currently a growing field in Nepal. Nepal has evolved over the years, with the introduction of new technologies and advanced imaging techniques. However, there are still challenges that need to be addressed, such as the shortage of trained radiologists. Only 300 radiologists are registered in Nepal as of 2021, which is insufficient to service the entire country's population, according to the Nepal Medical Council. In remote locations, where there are few or no radiologists accessible, the lack of radiologists is extremely severe, making it challenging for patients to get diagnostic imaging services. The availability of skilled radiologists is essential for the difficult process of identifying various chest ailments from X-rays.

This paper presents a model which can automatically detect 14 different chest conditions namely Effusion, Cardiomegaly, Emphysema, Nodule, Pneumonia, Pleural Thickening, Hernia, Fibrosis, Infiltration, Pneumothorax, Edema, Consolidation, Mass and Atelectasis from chest X-Rays. This paper proposes the use of customized Densent-169[1] for the classification of chest X-Rays. The paper proposes the use of a customized weighted cross-entropy loss to handle the problem of class imbalance in the dataset. This paper also proposes using Grad-CAM[1][2] to create images of each class with localized heatmaps.

The performance of this model was compared to CheXNeXt[3]. The training period of DCXNet was significantly less than CheXNeXt. CheXNeXt required nearly 20 hours for each training stage while DCXNet only required 4 hours of training. It was found that the presented model, DCXNet performs on par with CheXNeXt and even outperforms it on 4 of the chest conditions namely Cardiomegaly, Emphysema, Fibrosis and Hernia.

[1] Gradient-weighted Class Activation Maps

2. Literature Review

In the realm of medical image processing, categorization of chest X-rays has been a hotly debated subject. The authors of [1] defined a network architecture called Densenet. This architecture shows methods on how convolutional networks can be trained much more thoroughly, precisely, and quickly which helps in multi-label classifcation problems. The authors in [4] presented a dataset called ChestX-ray8, which contains 108,948 X-Ray images with 8 chest conditions. They also demonstrated that these 8 conditions can be located via weakly-supervised multi-label image classification. The authors of [5] introduced and assessed a partial resolution that utilizes LSTMs to exploit connections among target labels to predict 14 pathological patterns from chest X-ray images. They achieved leading performance results on the most extensive publicly accessible chest X-ray dataset from the NIH without any pre-training in 2018. In order to diagnose 14 distinct chest disorders, the authors of [3] compares the effectiveness of the CheXNeXt algorithm to that of professional radiologists. The outcomes demonstrate that the algorithm performed on par with radiologists and has the ability to support clinical judgment. This paper produces best outcomes on ChestX-Ray14 dataset with the help of ensemble methods but requires vey large training periods as training is done in two steps. In the 1st step nearly 10 different networks are trained and then few of the best

performing models are then selected to create the final model. This does produce better results but at high cost of training time.

3. Methodology

3.1 Block Diagram

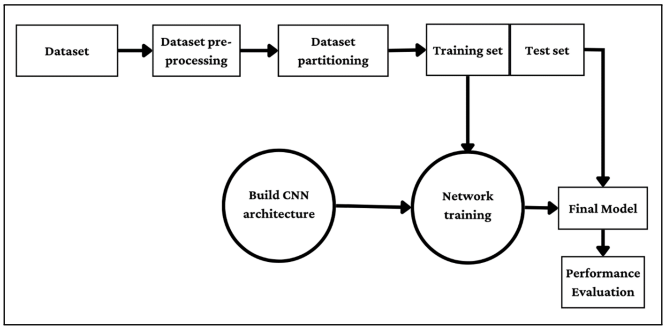

Figure 1: Block Diagram

3.2 Data

The data used for training the model was obtained from NIH[2] which contained 112,000 frontal-view X-Ray images of 30,805 unique patients. The dataset is known as ChestX-Ray14. The dataset contains 14 labels for each image with values either 0 indicating negative for the label and 1 meaning that the image is positive for the label.

3.2.1 Data Preprocessing

The total size of obtained database was 42.5 GB. The labels for all images were stored in a CSV file. The file was then loaded into dataframe using pandas. The path to each image was obtained and added to dataframe. All of labels were identified and one-hot-encoding was performed. The images were normalized based on the mean and standard deviation of images in the dataset. The images were then resized to target size of 320x320.

3.2.2 Training Set

The dataset was initially split randomly using group shuffle split into 70% and 30%, with 70% being the training set. Remaining 30% of data was then again split into two equal halves for test and validation sets. The training set contains 78,566 images while both the validation and test contains 16,777 images each. No data overlapping was found between these sets.

3.2.3 Test Set

After splitting the data, the test set had 16,777 frontal chest X-Rays. According to NIH, these images were annotated by four practicing radiologists at Stanford University. Radiologists were not given access to any patient information or were informed of the prevalence of any diseases in the data.

3.3 Loss function and Class Imbalance

The dataset that was used to train the model is prone to class imabalce problem. EDA[3] on the training dataset led to the following results:

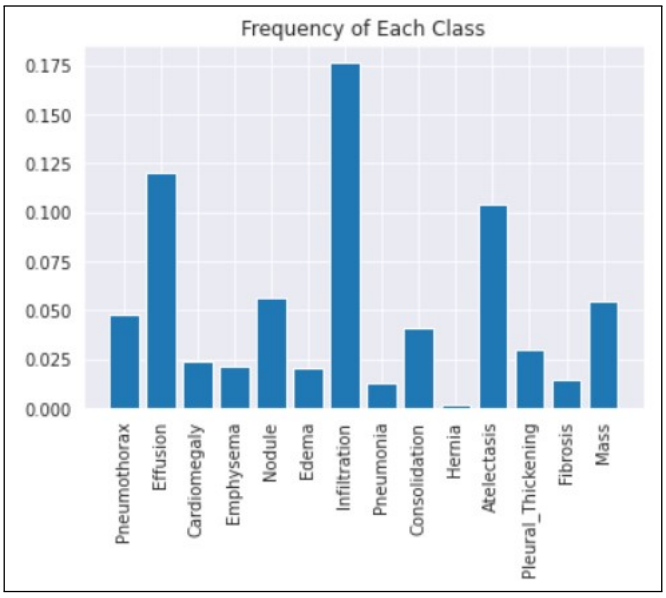

Figure 2: Imbalance in the dataset

- The most unbalanced pathology is Hernia, with 0.1% of patients testing positive for training..
- However, only 17.5% of the training instances for the Infiltration pathology, which has the least degree of imbalance, have been classified as positive.

This class imbalace issue doesnot allow for a normal cross-entropy loss for each class. For a balanced data set the loss function is:

$$\mathcal{L}(x_i) = -(y_i log(f(x_i)) + (1-y_i) log(1-f(x_i)))$$

where x_i and y_i are the input features and their corresponding labels and $f(x_i)$ is the output of the model which indicates the probability that it is positive. With the use of this formulation, we can observe that the loss will be dominated by the negative class in situations when there is a significant imbalance and there are few positive training events. One way of balancing such datasets require multiplying each class by a class-specific weight factors, w_p and w_n where w_p is the frequency of negative samples and w_n is the frequency of positive samples for each class. Then the previous unweighted loss function was modified as:

$$\mathcal{L}^w(x) = -(w_p y log(f(x)) + w_n(1-y) log(1-f(x)))$$

This equation was used to calculate the loss for each class and then the total loss can be calculated as the mean of the loss of each of the classes.

3.4 Model Architecture and Training

The presented model, DCXNet is a customized Densent-169[1], a 169 layer Convolutional Neural Network trained on the ChestX-ray14 dataset. DenseNets enhance information flow and gradients inside the network, making very deep network optimization feasible. The final fully connected layer was replaced with a global average pooling layer after which a fully

[2]National Institutes of Health
[3]Exploratory Data Analysis

connected layer which produces a 14-dimensional output was added. After which, element-wise sigmoid non-linearity was applied.

The weights of the presented model were initialized with weights from a model pretrained on ImageNet[6]. This network was trained using Adam[7] optimizer. Adam is a powerful variation of the stochastic gradient descent optimization technique, which iteratively adjusts parameters to reduce training-related loss. The model was trained in mini-batches of size of 32.

While training there were 100 steps per epoch and 50 validation steps for each epoch. Initial learning rate of $1e-4$ was used which was decayed by a factor of 20 each time the validation loss plateaued after 3 epochs. Model checkpointing was used to save the model every time after validation loss improved. Early stopping callback was used to stop the training once the validation loss didnot improve for the last 15 epochs. Other forms of regularization such as dropout or weight decay was not used. Then, the model with the lowest validation loss was picked. Each step of training completed after 3 minutes on a single NVIDIA Tesla P100. The presented network had 12,507,790 trainable parameters.

4. Results and comparision with previous State-of-the-Art models

The presented model outputs a vector for each label with values ranging between $0-1$ providing the probability of presence of the following 14 chest conditions: Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Emphysema, Fibrosis, Hernia, Mass, Nodule, Pleural Thickening, Infiltration, Pneumonia and Pneumothorax.

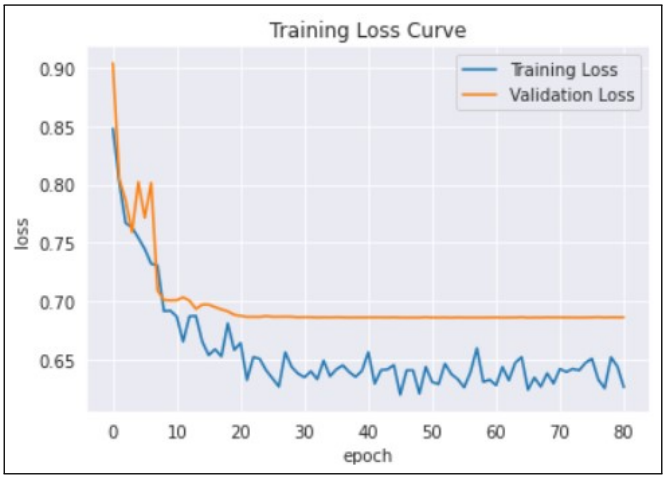

Figure 3: Training/Validation Loss Curve

Figure 3 shows the training/validation loss curve. In the image one can see that both training and validation losses are high in the beginning. They gradually decrease upto the 20th epoch. After which, the validation loss decreases very less with respect to training loss. However, the least validation loss is obtained at 65th epoch. The training was stopped at 80th epoch by early stopping callback. Spikes can be seen in the curve which is the effect of using mini-batch gradient descent.

To compare the DCXNet to previous algorithms, a single diagnostic performance measure, AUC[4] was used. AUC is a metric that indicates how well a model fits. In the medical field, this value also represents the likelihood that a patient who suffered a condition would have a greater risk score than a patient who didn't experience the event, if chosen at random. It summarizes the model's performance across various thresholds and provides a reliable indication of its ability to distinguish between different cases. AUC was chosen because it doesnot require a threshold value and we donot need to convert our model outputs to binary predictions as the model outputs a vector between $0-1$ for every class which indicates the probability of every class for a provided image. This AUC was calculated using sklearn library which plots a ROC[5] curve and then calculates the area under the curve using Simpson's rule.

The performance of the presented model was compared to Wang[4], Yao[5] and CheXNeXt[3] in Table 1. From the table, it can be seen that DCXNet performs on par with CheXNeXt on labels Atelectasis, Cardiomegaly, Effusion, Infiltration, Fibrosis, Hernia, Pnemothatorax and Pleural thickening (Difference of AUC ≤ 0.07). CheXNeXt has higher performance on Mass, Nodule and Pneumonia class(AUC ≥ 0.1). DCXNet has marginal improvement over CheXNeXt in Cardiomegaly, Fibrosis and Hernia class. DCXNet has larger improvement in the Emphysema class with the AUC difference > 0.2 .

One can also see that DCXNet performly poorly on Mass, Pneumonia and Nodule pathology if compared to CheXNeXt[3] but still better than Wang[4] and Yao[5]. This is because CheXNeXt uses ensemble methods which creates multiple models and chooses best among them to create a final model. This was a necessary trade-off to reduce the training period for DCXNet, hence DCXNet performs poorly on some classes than CheXNeXt.

Table 1: Performance comparision of presented model with state-of-the-art models on basis of AUC

Condition	Wang[4]	Yao[5]	CheXNeXt[3]	DCXNet
Atelectasis	0.716	0.772	0.862	0.804
Cardiomegaly	0.807	0.904	0.831	0.897
Effusion	0.784	0.859	0.901	0.874
Infiltration	0.609	0.695	0.721	0.706
Mass	0.706	0.792	0.909	0.792
Nodule	0.671	0.717	0.894	0.735
Pneumonia	0.633	0.713	0.851	0.735
Pneumothorax	0.806	0.841	0.944	0.876
Consolidation	0.708	0.788	0.893	0.823
Edema	0.835	0.882	0.924	0.89
Emphysema	0.815	0.829	0.704	0.907
Fibrosis	0.769	0.767	0.806	0.81
Plural Thickening	0.708	0.765	0.798	0.782
Hernia	0.767	0.914	0.851	0.856

5. Model Interpretation

Using class activation mappings(CAMs)[2], the presented model created heat maps to show the regions of the chest radiograph that contributed most to the network's categorization for use in interpreting predictions. To generate these CAMs, images were fed to the trained network and feature maps were obtained from final layer of the network. A map of most salient features

[4] Area Under the Curve
[5] Receiver Operating Characteristic

used to classify image containg a class is obtained by taking the weighted sum of feature maps using associated weights in the final fully connected layer. By scaling the map to fit the image's size and superimposing it on top of the image, the most crucial features, the model utilized to predict the condition were shown in the image.

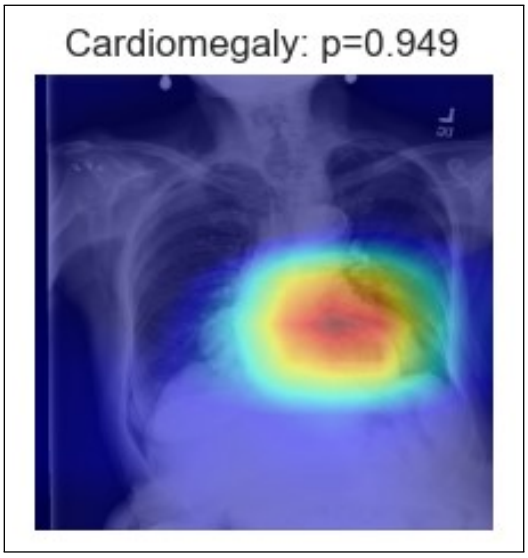

Figure 4: Patient with cardiomegaly. The model was able to correctly predict the presence of cardiomegaly (Enlarged Heart) and localize it using CAM

6. Conclusion

The majority of patient morbidity and mortality is caused by disorders of the chest. In order to avoid consequences, including death, early diagnosis of these illnesses is essential. The most often used imaging examination tool in practice is the Chest X-Ray, which is essential for the screening, diagnosis, and treatment of a number of disorders. However, the shortage of radiologists is particularly acute in rural areas in Nepal which makes it difficult for patients to access diagnostic imaging services. Even with the availability of imaging technology, there is a lack of specialists who can interpret X-rays, which increases mortality from diseases that are curable.

This paper presents a model, DCXNet, which detects 14 different chest conditions from frontal-view chest X-Ray images whose performance is on par with CheXNeXt[3] which claims their performances are at a level exceeding practicing radiologists. The presented model has a mean AUC of 0.82 while CheXNeXt has mean AUC of 0.84. This marginal improvement of CheXNeXt is the result of using ensemble methods which takes significantly larger training time. But the proposed model was able to achieve on par score with CheXNeXt just by replacing Densenet-121 in CheXNeXt algorithm with Densenet-169. The training time was significatly reduced only to 4 hours because of using transfer learning.

With the automation of this calibre, we anticipate that this technology will enhance the delivery of healthcare and provide access to medical imaging expertise in regions with a shortage of qualified radiologists.

7. Limitations

There are a few limitations of the presented model. Only frontal X-Rays were used to train this model but some of the conditions require lateral view for accurate diagnosis. Thus, it is anticipated that this configuration offers a conservative estimate of performance. The model's inability to access patient history is another drawback, which lowers its diagnostic performance when analyzing chest X-rays.

Acknowledgments

The authors would like to express their sincere gratitude towards the Department of Electronics and Computer Engineering, Purwanchal Campus for their invaluable contribution to this project. Their support, guidance, and expertise were instrumental in achieving the successful completion of this project. Finally, the authors would like to extend their appreciation to their colleagues and friends who provided them with their unwavering encouragement and support throughout the duration of this project.

References

[1] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. 2016.

[2] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. 2019.

[3] Pranav Rajpurkar, Jeremy Irvin, and Brandon Yang Robyn L. Ball, Kaylie Zhu. Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists. 2018.

[4] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. 2017.

[5] Li Yao, Eric Poblenz, Dmitry Dagunts, Ben Covington, Devon Bernard, and Kevin Lyman. Learning to diagnose from scratch by exploiting dependencies among labels. 2018.

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. 2009.

[7] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. 2014.