Domain Specific Deep Neural Network Model for Classification of Abnormalities on Chest Radiographs

Afolabi B. Samuel

Received: 1 January 1970 Accepted: 1 January 1970 Published: 1 January 1970

Abstract
This study collected, pre-processed dataset of chest radiographs, formulated a deep neural network model for detecting abnormalities. It also evaluated the performance of the formulated model and implemented a prototype of the formulated model. This was with the view to develop a deep neural network model to automatically classify abnormalities in chest radiographs. In order to achieve the overall purpose of this research, a large set of chest x-ray images were sourced for and collected from the CheXpert dataset, which is an online repository of annotated chest radiographs compiled by the Machine Learning Research group, Stanford University. The chest radiographs were preprocessed into a format that can be fed into a deep neural network. The preprocessing techniques used were standardization and normalization. The classification problem was formulated as a multi-label binary classification model, which used convolutional neural network architecture for making decision on whether an abnormality was present or not in the chest radiographs. The classification model was evaluated using specificity, sensitivity, and Area Under Curve (AUC) score as parameter. A prototype of the classification model was implemented using Keras Open source deep learning framework in Python Programming Language. The AUC ROC curve of the model was able to classify Atelestasis, Support devices, Pleural effusion, Pneumonia, A normal CXR (no finding), Pneumothorax, and Consolidation. However, Lung opacity and Cardiomegaly had probability out of less than 0.5 and thus were classified as absent. Precision, recall, and F1 score values were 0.78, this imply that the number of False Positive and False Negative are the same, revealing some measure of label imbalance in the dataset. The study concluded that the developed model is sufficient to classify abnormalities present in chest radiographs into present or absent.

Index terms — transfer learning, convolutional neural network, radiograph, classification, multi-label.

1 Introduction
Transfer Learning (Pan and Yang, 2009) is an important concept in machine learning research (Tan et al., 2018) that allows the domains, tasks, and distributions used in training and testing a network to be different from each other (Pan and Yang, 2010). Transfer learning is used to improve a learner from one domain by transferring information from a related domain (Weiss et al., 2016). The relatedness interacting domains cannot be over emphasized, as this could impact the relevance and appropriateness of the results.
that are proportional to the absorption characteristics of tissue with respect to a signal projected through the
body (Petrou and Petrou, 2010), care should be taken to employ cross-domain transfer learning. Even though,
transferring knowledge (learned features) from loosely related datasets such as ImageNet (Deng et al., 2009) to
medical image in situations where there is insufficient ground-truth label may be promising, but it may introduce
unintended biases which are undesirable in a clinical setting (Wang et al., 2018). As a result of the drawback
of transfer learning especially in a sensitive domain like medicine, a better alternative is to train deep learning
models exclusively of medical images. This is called training from scratch. Training a DL model from scratch is
not without computational bottlenecks. This is why research in this area is limited because of insufficient labeled
dataset (Tajbakhsh et al., 2016). However, advances in medical imaging technology and concern of deep learning
research community on medical image analysis has led to the production of more radiological certified annotation
on medical images. Notable among such is the National Institutes of Health (NIH) chest radiograph collection
that consists of more than 100,000 chest radiographs with annotations (Wang et al., 2017).

The Stanford University also released a very large dataset of chest radiographs with labels of pathology. This
dataset contains over 200,000 open sourced chest x-rays (Irvin et al., 2019).

Due to the availability of these dataset resources, one of the major reasons for cross-domain transfer learning
has been eliminated. Data augmentation methods can be employed to extrapolate the available medical images
to more than ten times the original quantity, so that more data could be used to learning medical image analysis
models. Data augmentation techniques would also be such that would not remove relevant information from the
medical images. Also, given that computational resources are available, then deep neural network models could
be trained from scratch on medical images; hence this study.

2 II.
3 Statement of Research Problem

cXR are often characterized by variability in contrast intensity and texture, which are different from the content
and structure of the images of natural object. Most of the existing deep models for medical image analysis are
based on transfer learning that depends largely on the fine tuning of feature weights learned from natural image
dataset. However, this cross-domain knowledge transfer is often not suitable for medical image analysis due to
its inability to handle the variability in contrast intensity and texture that characterizes medical images. Again,
in natural objects relative pixel intensity is used to convey information about a target object. That is, intensity
variation and saturation are irrelevant when handling natural images. In contrast, medical images use exact
pixel intensity values to convey information about abnormalities present in medical images. This intensity values
are represented using the Hounsfield scale (Prince and Links, 2006). Also, location invariance does not affect
the information content of natural images. In medical images on the other hand location is used to indicate
pathological sites, because certain abnormalities are more likely to appear in certain part of a scan or x-ray.
But the location of a dog or plate does not mean it is not a plate. Dimensionality reduction is a popular
technique used to enhance the performance and efficiency of deep models, natural objects are scale invariant that
is, and they retain their meaning irrespective of the scale. In medical images when the scales are change certain
information-reach contents tend to lost. Deep learning models and architectures are developed using natural
objects which still perform with relative intensity, location and scale. Therefore, the outcome of medical
image analysis from deep models that are trained using transfer learning mechanism are often not acceptable
by the medical professionals and as such implemented CAD system from these deep models are rarely deployed
for clinical practices. Therefore there is need to train medical image analysis models using medical image data;
hence this study aims at developing a Deep Convolutional Neural Network model using domainspecific data for
classification of abnormalities in chest x-rays.

4 III.
5 Objectives of the Research

The specific objectives are to acquire and preprocess dataset of chest radiographs; formulate a deep network model
for detecting abnormalities; evaluate the performance of the formulated model; and implement a prototype of
the formulated model IV.

6 Concept of Transfer Learning

The concept of transfer learning was motivated by the fact that people can always apply previous knowledge to
solve new problems faster (Torey and Shavlik, 2010) because repetition of common knowledge in the new task
is abstracted away. It is a learning approach where knowledge from a domain is applied to solve a problem in
another related domain. It is also referred to as domain adaptive learning (Kouw and Loog, 2019). Transfer
learning was inspired by the natural ability of human being to intelligently and intuitively apply knowledge from
previous task to tackle new and previously unseen task. In ML applications, algorithms are developed to solve
specific task such as classification, regression, or clustering problems. These algorithms often required labeled
dataset for good generalization. It is expected that the training data and test data are in same feature space
For this study, the CheXpert dataset was used.
9 a) Dataset Preparation and Preprocessing

To ease label matching to images, the uncertainty labels (-1.0) was converted to positive labels (1.0). This is to achieve a binary mapping of all labels similar with the U-ones model of Irvin et al. (2019). In statistic this is called zero imputation strategy (Kolesov et al. 2014). The assumption here is that diseases that are not sure to be present in the CXR (uncertainty label) could be coded as present. On the other hand, label categories that are referred to as unmentioned (blanks cells) were coded as negative (0.0) or absent. This approach follows the principle in literature which is known as zero imputation strategies.

10 Analysis of Result from the Developed Model

The entire dataset was not used for the model training because of the unexpected computational complexity and overhead. Attempt on training the model in the entire training set produced a memory error. As a result, the training was carried out on the 234 images of the validation set with 0.1 used as test case. The network was trained over 10 epoch, this means that the model iterated over the train dataset 10 times. The model summary is represented graphically as model loss, model accuracy and the AUR ROC curve. Also, after model training Precision, recall, F-score and the accuracy was given as output.

From Figure 3, the model accuracy showed that the accuracy increases rapidly in the first two epochs, indicating that the network is learning fast. Also, it showed that the model could probably be trained a little more as the trend for accuracy on both datasets is still rising for the last few epochs. Again, it is seen that the model has not yet over-trained the learning dataset.

From the Model loss curve as shown in Figure ??

11 AUC ROC Analysis of Result

The AUC ROC curve presented in Figure ?? was used to visualize all the performance metrics from a single image (a test image). The ROC curve was plotted for all the abnormalities present in the CXR of the CheXpert dataset. Table 3 captures the recorded probability values from the AUC ROC curve. From the AUC ROC curve, the developed model discriminated against some abnormality. Using the user-defined threshold value of 0.5, the threshold value was got from the average of the probability outcome of 0.57. Therefore, the developed model confidently detected the presence Atelestasis, Support devices, Pleural effusion, Pneumonia, A normal CXR (no finding), Pneumothorax, and Consolidation. This happened because the probability score was beyond the set threshold. However the model was detected the following abnormality as absence using the user defined threshold value of 0.5. That is, the probability outputs of Lung opacity and Cardiomegaly were less than 0.5. The developed model was not able to detect lung lesion, pleural other and fracture therefore a NAN value was returned. The accuracy of the model was approximately 0.78. This means the model can predict the presence of an abnormality 78 times in a given 100 cases. Precision and recall had the same have value that is 0.78.

12 IX.

13 Conclusion

This research has developed a model with medical images using Convolutional Neural Network. This is called training from the scratch because it does not involve the use of pre-trained weights. The developed model was based on the binary loss function because the problem was reduced to a binary multilabel problem where the model could detect the presence (1) or absence (0) of abnormalities under consideration. Fourteen ( 14) abnormalities associated with chest x-rays were examined in this study. The choice of model parameter was hampered by the limitation of computational resources. Hence, the only parameter that was tweaked was the batch size and image dimension. The two parameters were adjusted to optimally utilize the available computational resources. Image was down-sized to 250 * 240 and the batch size was changed often time but was pegged to 6 at the last computation stage. The number of network layer was only five, namely, the input layer, convolution layer, pooling layer and the fully connected layer. The model favorably predicted some abnormalities such as pneumonia, consolidation, pleural effusion, normal chest x-rays, pneumothorax. On the other hand abnormalities such as lung opacity and cardiomegaly were not well predicted by the model. Model returned a null value for lung lesion, pleural other and fracture due the presence of uncertainty label. To handle label cooccurrence the model threshold was used to determine abnormalities that have likelihood of co-occurring in a chest x-ray.

14 X.

15 Limitation of the Work and Future Research Direction

The research is however faced with a lot of issue on computational resource. Training model from scratch is highly computationally intensive. This was a serious limitation to getting the desired performance. Also a lot of parameter that was supposed to be tweaked was not done. The model has only four (4) Convolution layers which supposed to be deeper. Data augmentation was not carried out because of the processing capacity of the machine and the dimension of the input data. Care was taken not to downsize the image size more that necessary to avoid image
degradation and loss of details from the chest x-ray. Transfer learning is a fast and quick technique for developing deep learning model, but on medical images it weight must be from domain similar to medicine. Also, more computational resources like High Performing Computers (HPC) so that massive data could be used.

Figure 1:
LIMITATION OF THE WORK AND FUTURE RESEARCH DIRECTION

Figure 2: Figure 1:

<table>
<thead>
<tr>
<th>Pathology</th>
<th>Positive (%)</th>
<th>Uncertain (%)</th>
<th>Negative (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Finding</td>
<td>16627 (8.86)</td>
<td>0 (0.0)</td>
<td>171014 (91.14)</td>
</tr>
<tr>
<td>Enlarged Cardiom.</td>
<td>9020 (4.81)</td>
<td>10148 (5.41)</td>
<td>168473 (89.78)</td>
</tr>
<tr>
<td>Cardiomegaly</td>
<td>23002 (12.26)</td>
<td>6597 (3.52)</td>
<td>158042 (84.23)</td>
</tr>
<tr>
<td>Lung Lesion</td>
<td>6856 (3.65)</td>
<td>1071 (0.57)</td>
<td>179714 (95.78)</td>
</tr>
<tr>
<td>Lung Opacity</td>
<td>92669 (49.39)</td>
<td>4341 (2.31)</td>
<td>90631 (48.3)</td>
</tr>
<tr>
<td>Edema</td>
<td>48905 (25.06)</td>
<td>11571 (6.17)</td>
<td>127165 (67.77)</td>
</tr>
<tr>
<td>Consolidation</td>
<td>12730 (6.78)</td>
<td>23976 (12.78)</td>
<td>150035 (80.44)</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>4576 (2.44)</td>
<td>15658 (8.34)</td>
<td>167407 (89.22)</td>
</tr>
<tr>
<td>Atelectasis</td>
<td>29333 (15.63)</td>
<td>29377 (15.66)</td>
<td>128931 (68.71)</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>17313 (9.23)</td>
<td>2663 (1.42)</td>
<td>167665 (89.35)</td>
</tr>
<tr>
<td>Pleural Effusion</td>
<td>75696 (40.34)</td>
<td>9419 (5.02)</td>
<td>102526 (54.64)</td>
</tr>
<tr>
<td>Pleural Other</td>
<td>2441 (1.3)</td>
<td>1771 (0.94)</td>
<td>183429 (97.76)</td>
</tr>
<tr>
<td>Fracture</td>
<td>7270 (3.87)</td>
<td>484 (0.26)</td>
<td>179887 (95.87)</td>
</tr>
<tr>
<td>Support Devices</td>
<td>105831 (56.4)</td>
<td>898 (0.48)</td>
<td>80912 (43.12)</td>
</tr>
</tbody>
</table>

Figure 3: Figure 2:
Figure 4:
LIMITATION OF THE WORK AND FUTURE RESEARCH DIRECTION

Figure 5: Figure 3

Figure 6: Figure 4 :Figure 5 :
It was compiled by the Stanford University Machine Learning Group. The dataset contains 224,316 chest radiograph images of 65,240 patients. It consists of fourteen radiological observations with consideration for uncertainty in radiograph interpretation.

Figure 7:

[Note: Source: Irvin et al, 2019]

Figure 8: Table 1 :

<table>
<thead>
<tr>
<th>Initialized Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch size</td>
</tr>
<tr>
<td>Initial Learning Rate</td>
</tr>
<tr>
<td>Epoch</td>
</tr>
<tr>
<td>Epsilon</td>
</tr>
<tr>
<td>Kernel size</td>
</tr>
</tbody>
</table>

VII.

Figure 9: Table 2 :

<table>
<thead>
<tr>
<th>Year</th>
<th>Volume</th>
<th>Issue</th>
<th>Version</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023</td>
<td>XXIII</td>
<td>I</td>
<td>I</td>
<td></td>
</tr>
</tbody>
</table>

( ) D

<table>
<thead>
<tr>
<th>Features (abnormality)</th>
<th>Output probabilities</th>
<th>Global Journal of Computer Science and Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enlarged Cardiomegaly</td>
<td>0.63 0.54 0.47</td>
<td></td>
</tr>
<tr>
<td>Lung opacity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lung lesion</td>
<td>0.44 Nan 0.50</td>
<td></td>
</tr>
<tr>
<td>Edema</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>Consolidation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pneumonia</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>Atelectasis</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>Pleural effusion</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Pleural Other</td>
<td>Nan</td>
<td></td>
</tr>
<tr>
<td>Fracture</td>
<td>Nan</td>
<td></td>
</tr>
<tr>
<td>Support Devices</td>
<td>0.65</td>
<td></td>
</tr>
</tbody>
</table>

Figure 10: Table 3 :


[Medical Tasks. in CLEF (Working Notes)] Medical Tasks. in CLEF (Working Notes),


11