Detecting Types of Pneumonia in NIH Chest X-Rays using Machine Learning and Deep Learning

Arsh Irfan Modak, Omkar Shivaji Waghmare, Adithya Chenthilkannan

1. Problem Description

What problem are you solving?

Chest X-ray exams are one of the most frequent and cost-effective medical imaging examinations available. However, clinical diagnosis of a chest X-ray can be challenging and sometimes more difficult than diagnosis via chest CT imaging. By using machine learning and deep learning techniques, we aim to ease the detection and classification of pneumonia and its types.

Describe the problem formally from a computational perspective.

The project will be implemented in Python using various libraries like Pandas, NumPy, PyTorch etc.

<u>Initial State:</u> Untrained models
<u>Transition Function:</u> Training of the models, backpropagation of the weights and biases, loss computation based on loss function.
<u>Goal State:</u> convergence of algorithm or end of training via maximum iterations/number of epochs.

Initially, before training, all deep learning models will have 0 weight and bias - once we start training, these weights and biases will be updated based on the loss function, optimizer and learning rate. Once the entire dataset has been passed through the model training phase once, one epoch is completed. We aim to train our models for at least 50 epochs. Once fully trained, the learned architecture along with its weights and biases will return a list of classes with their corresponding probability for a particular input image.

For traditional models, depending on the algorithm, a decision boundary will be learned to separate the 15 distinct classes.

What are the inputs and outputs (exactly)?

1. <u>For Deep Learning:</u> The input will be 224x224x3 images and the output will be the 15 distinct classes along with their probabilities.
2. <u>For Supervised Machine Learning:</u> The input will be the extracted features from the CNN or other techniques (size will differ on architecture and/or technique) and the output will be the 15 distinct classes along with their probabilities.

What data are you using (exactly)?

Our dataset consists of 112,120 1024x1024 single band X-ray images with disease labels from 30,805 unique patients. We also have a CSV file with the image names and the labels. The labels are expected to be >90% accurate.

Why is it interesting?

Working with images involving the healthcare domain is specifically interesting to the team and the solution makes the most social impact. In our case using Supervised machine learning is unconventional to the problem and making it work is an interesting task.
Working with such a large amount of data was something the three of us wanted to experience, hence, managing and utilizing the entire data with limited resources is also an interesting aspect of this project.

2. Algorithms

What algorithms do you use?

1. Supervised Machine Learning:
   1.1 Linear Model: Multinomial Logistic Regression (Arsh Modak)
   1.2 Non-Linear Model: SVM with non-linear kernels (Omkar Waghmare)
   1.3 Ensemble Model: Random Forest (Adithya Chenthilkannan)

2. Deep Learning:
   2.1 VGG16 (Adithya Chenthilkannan)
   2.2 Densenets (Omkar Waghmare)
   2.3 Resnets (Arsh Modak)

3. Baseline Model: we will create a baseline CNN for the sake of comparisons with SOTA CNN architectures.

How are these algorithms typically used, and how are you using them?
Typically traditional supervised machine learning algorithms which are used for classification take training and validation data as the input with the target variable as the output and learns relations, patterns between the input and output variables, For our case, instead of feeding images to these models directly we will use multiple feature extraction techniques (CNN's, Harlick Textures and HU moments) and feed these features into the models along with their target variables.

Similarly Deep Learning algorithms also predict/classify the target variables but incorporate non-linearity and flexibility in their decision making process. Additionally NNs are great feature extractors and learn relationships, identify patterns within the data during training, hence, we will be directly feeding our images with minimal transformations to them.

Why are these algorithms appropriate?
Having a well labelled dataset allows us to use supervised learning techniques such as supervised machine learning and deep learning algorithms. Although supervised machine learning algorithms are not the ideal choice while working with images, they have their own benefits (explainable, simple,, less resource intensive etc) and can be used as replacements if comparable results are achieved, hence, we believe traditional machine learning algorithms are appropriate for our use case. Our aim is to compare these traditional models with popular SOTA CNN architectures.
CNN architectures are popular for achieving outstanding results in image classification tasks such as ours, hence we will be experimenting with various architectures of CNNs.

Have other people used similar algorithms to solve your problem before?
Typically, for image classification of chest x-rays, deep learning models like SOTA CNN architectures are used. However, we did not find substantial work involving traditional machine learning algorithms.

3. Results

What results do you expect to show?
Sample images will be fed into the models and the predicted label will be returned as the output. We will create a grid of sample images along with their actual and predicted labels. The train and validation loss of each model. A bar plot of probabilities of all classes for a sample image.

What comparisons will you do?
All models that we have implemented will be compared with each other. Results of the various feature extraction techniques for the traditional models will also be included in the comparison.
Evaluation metrics such as Precision, Recall, F1 Score, Confusion Matrix, Accuracy, etc will be used to compare the results.
We will also compare the time it takes to train these models, and their losses.

Are there risks for not getting all the results? If so, what will you do about it?
Due to the size of the dataset, the models may take way too long to train. To overcome this, we may reduce the size of the data by working with a balanced sample of the dataset.

4. References

1. https://www.kaggle.com/nih-chest-xrays/data
2. https://nihcc.app.box.com/v/ChestXray-NIHCC/folder/36938765345
3. https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community
4. https://github.com/paloukari/NIH-Chest-X-rays-Classification#Technologies