Machine Learning Engineer Nanodegree

Capstone Proposal

Nory Diankov
June 18, 2018

Proposal

Domain Background

Much recent interest has been garnered in the application of AI in healthcare, particularly in the diagnosis of diseases. Lung conditions, a category that includes both communicable and non-infectious diseases have remained the second leading cause of death globally in the last 15 years, just after heart disease, according to the World Health Organization. In 2016 alone, diseases such as chronic obstructive pulmonary disease, tuberculosis, lung cancer, and lower respiratory infections accounted for 9 million deaths (The World Health Organization , 2018).

Despite the critical need for early screening and detection, it is estimated that two-thirds of countries do not have sufficient access to basic radiology services, such as a simple x-ray or ultrasounds. In particular, low-income countries are handicapped by an insufficient infrastructure and a considerable burden of disease, compounded by the need to allocate scarce resources to basic necessities such as clean water and nutrition. As a result, common limitation to the high mortality rate is a lack of staff and the cost of hiring radiologists (Silverstein, 2016).

Problem Statement

A lack of access to well-trained radiologists could delay diagnosis and the prevention and identification of deadly lung diseases. It is estimated that in a country of 43 million, Kenya has only 200 radiologists, whereas, one Boston hospital, Massachusetts General, alone has 126. Although there have been experiments in telemedicine—the practice of available experts in the US and Canada reading and diagnosing electronic medicine records of patients in countries of high-need—there are numerous challenges to overcoming delays associated with different time zones and the speediness of response (Wamala, 2013).

Therefore, the problem is to develop software that can 1) distinguish between normal and abnormal x-ray images, and 2) perform “diagnosis”, which in the case of radiological evidence usually entails implicating several possible conditions of

roughly equal probability. This study aims to begin the development of such a tool; for the purposes of the Udacity nanodegree, my goal is to show an algorithm that is capable of these tasks, even if it is not fully optimized.

Datasets and Inputs

I will use a dataset of 5,232 chest X-ray images taken from 5,856 pediatric patients, 1 to 5 years old, from the Gaungzhou Women's and Children's Hospital. Academic physicians have classified 3,883 of these as depicting pneumonia, within which 2,538 bacterial and 1,345 viral pneumonia cases, and 1,349 images as normal. The dataset is publicly available (Kermany, Zhang, & Goldbaum, Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification) and was used in a published study (Kermany & al., 2018). The images are of varying sizes and are grayscale.

Solution Statement

There is a substantial public-health need, especially in developing countries that lack robust medical infrastructure, for the development of accurate diagnostic imaging tools for the diagnosis of lung diseases. Such tools will alleviate some of the economic burdens of training qualified radiologists and can save lives in countries where even common lower respiratory infections are a substantial source of mortality due to the lack of timely diagnosis.

This study seeks to utilize deep learning to develop a model that is then trained on this database to identify and classify abnormalities in chest x-ray images.

Benchmark Model

Recent forays into CAD regarding pulmonary diseases have been made by researchers, particularly at Stanford University. A study published in 2017 utilized an algorithm, CheXNet, of a 121-layer convolutional neural network that takes X-ray image and returns an output of the probability of pathology. In this specific case, CheXNet focuses on identifying pneumonia, with a F1 performance of 0.435 that exceeds the average radiologist performance of 0.387 (Rajpurkar, et al., 2017).

The model architecture of CheXNet utilizes DenseNets to the improve flow of information and gradients through the network, with the final fully connected layer replaced with one that has a single output. A sigmoid nonlinearity was applied thereafter. Network weights were initialized with those from a model pretrained on ImageNet, using Adam with standard parameters ($\beta 1 = 0.9$ and $\beta 2 = 0.999$) and minibatches of size 16. The authors used an initial learning rate of 0.001 that is decayed by a factor of 10 each time the validation loss plateaus after an epoch, and pick the model with the lowest validation loss.

The dataset that I am using also formed the basis of a study published in Cell in 2018. This study used a basic, ImageNet-pretrained network that was directly applied to the images with very minimal changes to account for the number of labels in the classifier.

Evaluation Metrics

To measure the effectiveness of the proposed model's performance, the solution will be compared to this benchmark model using the published F1 score of CheXNet and the F1 score average performance of radiologists. The F1 score is the harmonic average of the precision and recall of the models, often used in the field of information retrieval for measuring classification performance (Powers, 2011), as shown visually below (F1 Score).

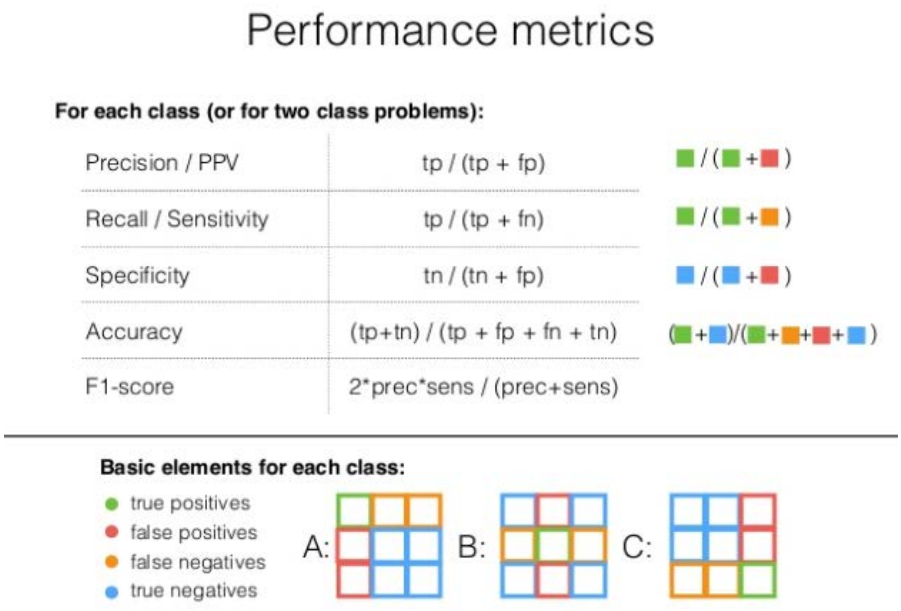

Fig. 1 Performance metrics, including F1, relevant to classification problems.

I will also use macro-averaging: averaging the performances of each individual class, where precision (PRE) over k-number of cases is defined as (Raschka):

$$PRE_{macro} = \frac{PRE_1 + \cdots + PRE_k}{k}$$

Project Design

I intend to build upon a well-known convolutional neural network, such as VGG-16, ResNet or DenseNet variants.

One approach I will try is to use multiple convolutional blocks where each consists of two or more conv2d layers of size 32 or 64 with kernels of 3x3 and nonlinear activation, a 2d-pooling layer and a dropout layer. This should enable multiple levels of features to be extracted and mapped. Finally, a global pooling may have to be applied and then a sequence of two fully-connected DenseNet layers with the final one using softmax, which will allow the direct interpretation of the resulting output weights in the output vector as probabilities. A topk method can finally retrieve the top-k most likely labels.

Another approach that I will try is to use the above, but with a residual network such as ResNet50. Residuals are attractive because “skipping” layers and connecting a layer’s output in a non-sequential manner has been shown to outperform deep networks with many more convolutional layers. Moreover, residual networks do not appear to suffer from the problem of saturating and even increasing errors with depth that is frequently seen in very deep networks with sequential linking.

I expect that I will have to perform the normal amount of data preprocessing and augmentation on the images. Since the networks I will build on were designed for ImageNet, most have fairly standard image input requirements such as size of 224x224 pixels, normalization of intensity between 0 and 1, and center-cropping. I do expect, however, that some more pre-treatment specific to the nature of X-ray images may have to be performed.

I plan to use PyTorch rather than Tensorflow due to the ease with which I can integrate Torchvision with my GPU.

Reference

F1 Score. (n.d.). Retrieved from https://www.slideshare.net/ThomasPloetz/bridging-the-gap-machine-learning-for-ubiquitous-computing-evaluation

Kermany, D., & al., e. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell, 172*(5), pp. 1122-1131.

Kermany, D., Zhang, K., & Goldbaum, M. (n.d.). Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification. Retrieved from https://data.mendeley.com/datasets/rscbjbr9sj/2

Powers, D. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 37-63.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Lungren, M., & Ng, A. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv:1711.05225v3*.

Raschka, S. (n.d.). Macro-averaging. Retrieved from https://sebastianraschka.com/faq/docs/multiclass-metric.html

Silverstein, J. (2016, September 27). Most of the World Doesn't Have Access to X-Rays. *The Atlantic*.

The World Health Organization . (2018). *Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016.* Geneva.

Wamala, D. S. (2013). A meta-analysis of telemedicine success in Africa. .*Journal of Pathology Informatics, 4*(6). doi:http://doi.org/10.4103/2153-3539.112686

Wang, X., Peng, Y., L, L., Z, L., M, B., & RM, S. (2017). ChestX-ray: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *IEEE CVPR*.