Synopsis

On

Chest disease detection from human X-ray scans using Deep Learning

Submitted in partial fulfillment of the requirement

For the award of the degree of

B.TECH

In

Computer Science & Engineering

Submitted By

Ravi Soni (1606810258) **Rohan Singh (1606810271)**

Shadib Shah (1606810285) **Shivang Malik (1606810296)**

Under the guidance

Ms. Sonika Jalhotra

Department of Computer Science & Engineering

Meerut Institute of Engineering & Technology, Meerut

Problem Statement:

The enormous number of chest radiographs produced globally are currently analyzed almost entirely through visual inspection on a slice-by-slice basis. This requires a high degree of skill and concentration, and is time-consuming, expensive, prone to operator bias (data distortion or wrong interpretation), and unable to exploit the invaluable informatics contained in such large-scale data.

Errors and delay in radiological diagnosis still contribute to 1 in 10 patient deaths in hospitals, making medical errors the third largest cause of death after heart disease and cancer.

Moreover, due to the complexity of chest radiographs, it is challenging even for radiologists to discriminate thorax diseases on them, resulting in the shortage of expert radiologists, particularly in rural areas who are competent to read chest radiographs.

Therefore, it is of significance to develop automated algorithms for computer-aided diagnosis of diseases on chest radiography.

Motivation behind the project:

Deep learning has transformed healthcare. It's being used to diagnose lung cancer, pneumonia, and other diseases. Deep learning is more accurate and faster at diagnosis than real doctors. Automation of X-ray analysis can prevent a lot of mishaps, speed up diagnosis and reveal new patterns thus aiding in medical research.

Objectives:

To implement a deep learning model using Deep Convolutional Neural Networks (DCNN) architecture which can predict various chest diseases like Pneumonia, Pneumothorax, Atelectasis, Effusion etc with significant accuracy (>80%) and provide other insights about the analysis performed by generating heat maps and other visualizations. The model will also be able to localize the pathology by generating Class Activation Maps(CAM). We will start first with Pneumothorax owing to the availability of good dataset.

Extended goals: Most of the deep learning techniques for disease detection can only find what they're looking for. If a learning model is looking for lung cancer, it's never going to find pneumonia or pneumothorax. To overcome this we will extend our model to incorporate Variational Autoencoders (VAE). To tackle the problem of limited availability of good datasets, Generative Adversarial Networks (GANs) will be used.

Scope:

A successful implementation of the project can be used to create a web application acting as a virtual radiology assistant. Various refinements will transform this implementation into a full blown service that can be used by experts in the medical field. These experts can than implement more features into the service which can then be used as a data sharing platform and as a training ground for medical students.

Methodology:

So how are human doctors today able to detect rare or new diseases in x-ray scans? A lot of the time they don't, but when they do it's because they find an anomaly. The doctor knows what a healthy lung looks like, and is able to notice when something doesn't look quite right. And that is in essence how the proposed implementation will work:

Training

1. The ChestXRay dataset consists of over 100,000 different X-rays from more than 30,000 patients labeled with different disease conditions. Let's start with one particular disease - Pneumothorax.
2. Pneumothorax is a condition that occurs when there are abnormal amounts of air in the space between the lung and chest wall. This reduces the capacity to which the lung is able to expand and fill with air, leading to oxygen shortage and low blood pressure. Lack of treatment may lead to worsening symptoms and even death.
3. Consider one training sample consisting of one healthy lung image and one with Pneumothorax.
4. Downscale the image samples to 256*256 pixels.
5. Remove noisy data (images with various illuminations, skewed images etc).
6. Configure the pretrained ResNet model to apply transfer learning.

Evaluation

1. TO evaulate our network, we use the **ROC** score. In a ROC curve, we plot the positive predictive value (true positives / (true positives + false positives)) against the negative predictive value (true negatives / (true negatives + false negatives)). We can get different positive and negative prediction values by choosing a different confidence threshold at which we consider a disease to be positive or negative. The area under the curve is then used as a metric to evaluate the quality of a model.

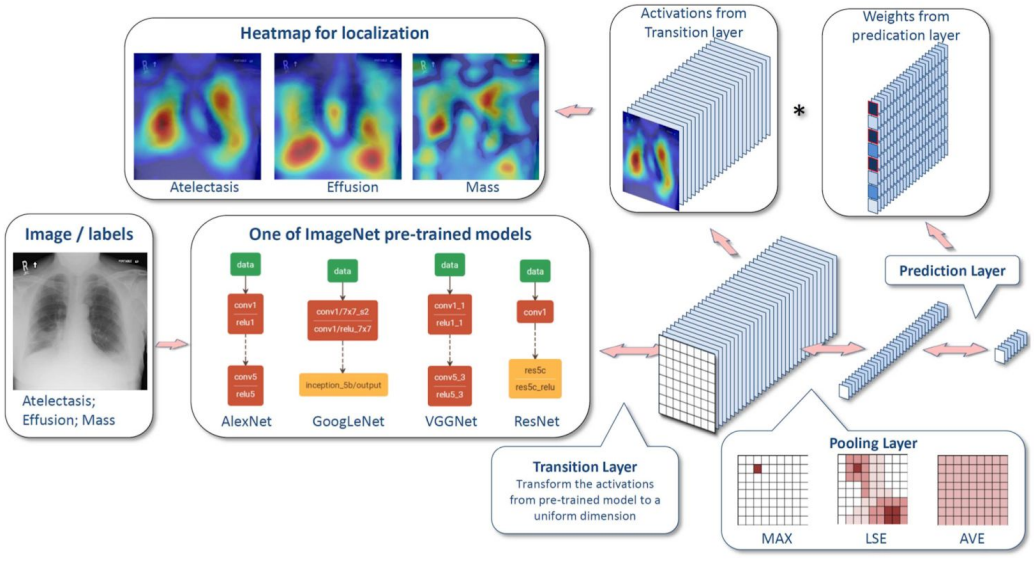
Interpretation of Model's output using Class Activation Mapping (CAM)

1. Class activation mapping generates "heat maps" indicating the regions of the image to which our model "attends" in the final layers.
2. Extract the appropriate feature map from an input image.
3. Compute the CAM for a few sample images. In order to visualize the relevant image regions, we'll need to upsample our CAM to our image size.

<u>Using GANs to learn what makes an image healthy</u>

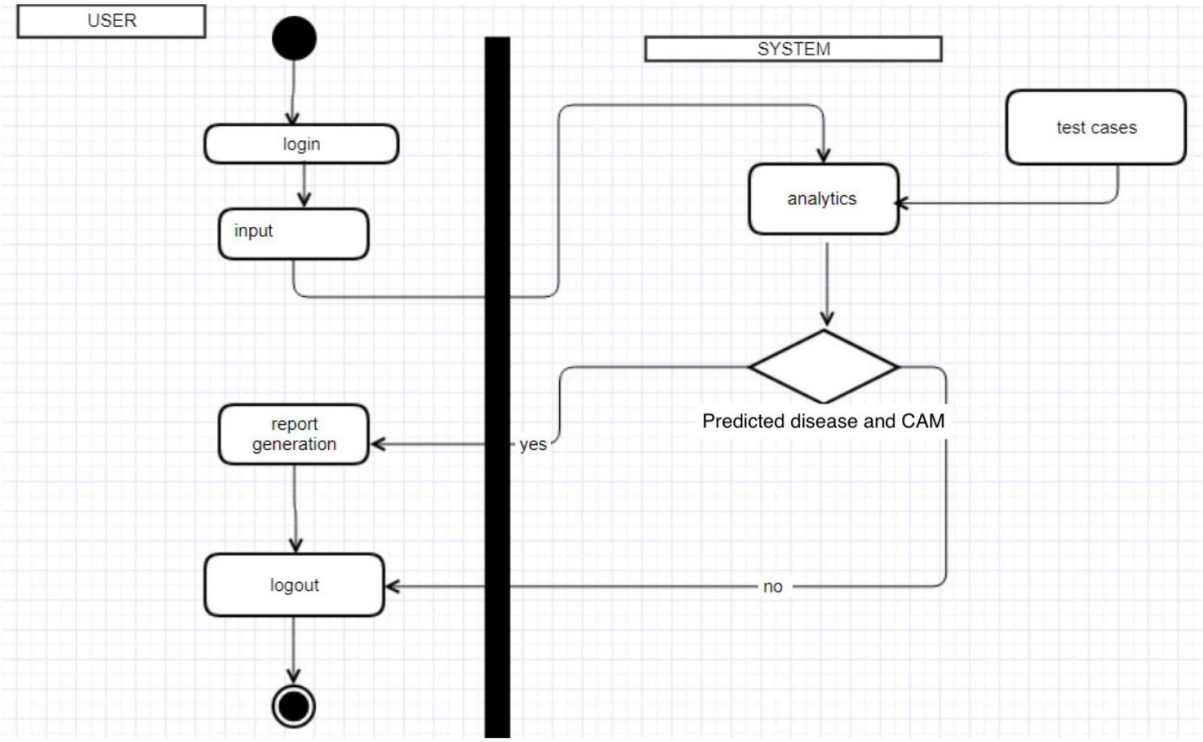
- In GANs, two neural networks are trained against each other. The first neural network is the Generator, which tries to generate ‘fake’ images of healthy lungs that look just like the real thing. The second neural network is the Discriminator, which tries to identify whether an image is real, or a fake created by the Generator.
- Since the GAN and autoencoder are only trained on images of healthy lungs when the autoencoder is given an image of an unhealthy lung, it’s reconstruction (output of the decoder) will look really bad. where the reconstruction is worst, is also where the disease is most prevalent. The diseased parts of the image are really different from what the autoencoder was trained on, and so the reconstruction will look most different from the original where the disease is.
- Using mean pixel difference, an algorithm that sees how different two images are, Zilic automatically highlights where the disease is in the image.
- We can also compare the encoding of the test image with the distribution of healthy images. If the encoding is within the distribution, the test image is healthy. If it’s outside of the distribution it’s unhealthy.
- Finally, we can use the discriminator from the GAN. It’s been trained to give low scores for healthy images of lungs, and it’ll give slightly higher scores for images of lungs with diseases.

Process Description:



1. Overall flow chart for DCNN and disease localization

We will use pre-trained ResNet model to speed up training.



2. Use Case

Hardware & Software:

Software:

1. Tensorflow and Pytorch for implementing the model and it's visualisation
2. Tensorboard for analysis
3. Flask for the backend
4. Jupyter notebook

Hardware: Owing to huge sizes of the dataset and complexity of the model, a GPU powered machine is required which supports CUDA in particular. For this purpose cloud service providers like GCP or Microsoft Azure can be used.

References:

1. Open-i: An open access biomedical search engine. [https: //openi.nlm.nih.gov]
2. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases [http://openaccess.thecvf.com/content_cvpr_2017/papers/Wang_ChestX-ray8_Hospital-Scale_Chest_CVPR_2017_paper.pdf]
3. Zhe Li, Chong Wang, Mei Han, Yuan Xue, Wei Wei, Li-Jia Li, Li Fei-Fei ., et al Thoracic Disease Identification and Localization with Limited Supervision. [https://arxiv.org/abs/1711.06373]
4. Li Yao, Eric Poblenz, Dmitry Dagunts, Ben Covington, Devon Bernard, Kevin Lyman., et al Learning to diagnose from scratch by exploiting dependencies among labels. [https://arxiv.org/abs/1710.10501].
5. X-ray datasets.
    a. Daniel Kermany, Kang Zhang, Michael Goldbaum https://data.mendeley.com/datasets/rscbjbr9sj/2
    b. CXR8 [https://nihcc.app.box.com/v/ChestXray-NIHCC]