Pnemonia & Data Science

Abstract:
machine learning techniques is very much active area of reasearch in medical science. With increasing size and complexity of medical data like X-rays deep learning gained huge success in prediction of fatal diseases like pnemonia.

Pneumonia is an inflammatory condition of the lung affecting primarily the small air sacs known as alveoli. Typically, symptoms include some combination of productive or dry cough, chest pain, fever and difficulty breathing.

History:

Pneumonia was regarded by Canadian pathologist William Osler in the 19th century as "the captain of the men of death".[16] With the introduction of antibiotics and vaccines in the 20th century, survival greatly improved.[11] Nevertheless, in developing countries, and also among the very old, the very young and the chronically ill, pneumonia remains a leading cause of death.[11][17] Pneumonia often shortens suffering among those already close to death and has thus been called "the old man's friend"

Pneumonia is a common illness affecting approximately 450 million people a year and occurring in all parts of the world.[11] It is a major cause of death among all age groups resulting in 4 million deaths (7% of the world's total death) yearly.[11][12] Rates are greatest in children less than five, and adults older than 75 years.[11] It occurs about five times more frequently in the developing world than in the developed world.[11] Viral pneumonia accounts for about 200 million cases.[11] In the United States, as of 2009, pneumonia is the 8th leading cause of death.[21]

In 2008, pneumonia occurred in approximately 156 million children (151 million in the developing world and 5 million in the developed world).[11] In 2010, it resulted in 1.3 million deaths, or 18% of all deaths in those under five years, of which 95% occurred in the developing world.[11][20][129] Countries with the greatest burden of disease include India (43 million), China (21 million) and Pakistan (10 million).[130] It is the leading cause of death among children in low income countries.[11][12] Many of these deaths occur in the newborn period. The World Health Organization estimates that one in three newborn infant deaths is due to pneumonia.[131] Approximately half of these deaths can be prevented, as they are caused by the bacteria for which an effective vaccine is available.[132] In 2011, pneumonia was the most common reason for admission to the hospital after an emergency department visit in the U.S. for infants and children.[133]

Cause :

Pneumonia is usually caused by infection with viruses or bacteria and less commonly by other microorganisms, certain medications or conditions such as autoimmune diseases
the disease may be classified by where it was acquired, such as community- or hospital-acquired or health care-associated pneumonia

Pneumonia is due to infections caused primarily by bacteria or viruses and less commonly by fungi and parasites. Although there are over 100 strains of infectious agents identified, only a few are responsible for the majority of the cases. Mixed infections with both viruses and bacteria may occur in roughly 45% of infections in children and 15% of infections in adults.[11] A causative agent may not be isolated in approximately half of cases despite careful testing

Diagnosis:

Symptom	Frequency
Cough	79–91%
Fatigue	90%
Fever	71–75%
Shortness of breath	67–75%
Sputum	60–65%
Chest pain	39–49%

Pneumonia is typically diagnosed based on a combination of physical signs and a chest X-ray

A chest radiograph is frequently used in diagnosis.[20] In people with mild disease, imaging is needed only in those with potential complications, those not having improved with treatment, or those in which the cause is uncertain.[20][60] If a person is sufficiently sick to require hospitalization, a chest radiograph is recommended.[60] Findings do not always match the severity of disease and do not reliably separate between bacterial infection and viral infection

Treatment:

Vaccines are available to prevent pneumonia caused by pneumococcal bacteria or the flu virus, or influenza

Role of Data science in detecting pnemonia:

China:

Infervision to gather medical data to train machine-learning algorithms in tasks like reading scans more easily than US or European rivals.

Infervision created its main product, software that flags possible lung problems on CT scans, using hundreds of thousands of lung images collected from major Chinese hospitals. The software is in use at hospitals in China, and being evaluated by clinics in Europe, and the US, primarily to detect potentially cancerous lung nodules.

Infervision's staff in Beijing worked through the Lunar New Year holiday to tune their existing pneumonia detection algorithms to look more specifically for Covid-19 pneumonia. The company acquired images of the newly discovered pneumonia from Wuhan Tongji Hospital, one of the first to receive patients with the new disease, and a longstanding collaborator. The version of the software in use today was trained with more than 2,000 images from Covid-19 patients

USA:

Flagler Hospital in Saint Augustine, Florida, is using artificial intelligence tools to improve the treatment of pneumonia.
The AI tools automatically revealed new, improved care pathways for pneumonia after analyzing thousands of patient records from the hospital and identifying the commonalities for those with the best outcomes

Ayasdi uses a branch of mathematics called **topological data analysis** to group patients treated similarly and find relationships between those groups. This analysis may result from AI in the form of supervised learning or unsupervised learning.

"Once the data loaded, they use unsupervised learning AI algorithm to generate treatment groups," "In the case of pneumonia patient data, Ayasdi produced nine

treatments groups. Each group was treated similarly and statistics were given to them to understand that group and how it differed from the other groups."

Need to investigate more on **topological data analysis.**

AI can be applied to various types of healthcare data (structured and unstructured). Popular AI techniques include machine learning methods for structured data, such as the classical support vector machine and neural network, and the modern deep learning, as well as natural language processing for unstructured data

convolutional neural network algorithm, utilizing a set of neurons to convolve on a given image and extract relevant features from them

Model from journal:(https://www.hindawi.com/journals/jhe/2019/4180949/)

Below is the overall architecture of the proposed CNN model which consists of two major parts: the feature extractors and a classifier (sigmoid activation function). Each layer in the feature extraction layer takes its immediate preceding layer's output as input, and its output is passed as an input to the succeeding layers. The proposed architecture in Figure 3 consists of the convolution, max-pooling, and classification layers combined together. The feature extractors comprise conv3 × 3, 32; conv3 × 3, 64; conv3 × 3, 128; conv3 × 3, 128, max-pooling layer of size 2 × 2, and a RELU activator between them. The output of the convolution and max-pooling operations are assembled into 2D planes called feature maps, and we obtained 198 × 198 × 32, 97 × 97 × 62, 46 × 64× 128, and 21 × 21 × 128 sizes of feature maps, respectively, for the convolution operations and 99 × 99 × 32, 48 × 48 × 64, 23 × 23 × 128, and 10 × 10 × 128 sizes of feature maps from the pooling operations, respectively, with an input of image of size 200 × 200 × 3 as shown in Table 2. It is worthy to note that each plane of a layer in the network was obtained by combining one or more planes of previous layers.

The classifier is placed at the far end of the proposed convolutional neural network (CNN) model. It is simply an artificial neural network (ANN) often referred to as a dense layer. This classifier requires individual features (vectors) to perform computations like any other classifier. Therefore, the output of the feature extractor (CNN part) is converted into a 1D feature vector for the classifiers. This process is known as flattening where the output of the convolution operation is flattened to

generate one lengthy feature vector for the dense layer to utilize in its final classification process. The classification layer contains a flattened layer, a dropout of size 0.5, two dense layers of size 512 and 1, respectively, a RELU between the two dense layers and a sigmoid activation function that performs the classification tasks.

The final results obtained are training loss = 0.1288, training accuracy = 0.9531, validation loss: 0.1835, and validation accuracy of 0.9373.