Pneumonia Detection Using Machine Learning

Premonvitha Sai Rayana

DATE: 23-10-2022

ABSTRACT:

Pneumonia is a respiratory infection caused by bacteria or viruses; it affects many individuals, especially in developing and underdeveloped nations, where high levels of pollution, unhygienic living conditions, and overcrowding are relatively common, together with inadequate medical infrastructure. Pneumonia causes pleural effusion, a condition in which fluids fill the lung, causing respiratory difficulty. Early diagnosis of pneumonia is crucial to ensure curative treatment and increase survival rates. Chest X-ray imaging is the most frequently used method for diagnosing pneumonia. However, the examination of chest X-rays is a challenging task and is prone to subjective variability. The medical establishment records that radiologist get it right on average only about 65 to 75% of the time and 5% of radiology interpretations contains clinically significant errors. Our AI algorithms analyse chest X-rays at both a higher level of accuracy and have faster report turnaround time than radiologists. This can be used to help radiologist with their workflow optimization . This will help patients obtain its early diagnosis and treatment in instance of abnormalities. Also a Noval Corona Virus pneumonia patient CT image is around 300 which brings greater pressure to the doctors clinical diagnosis. The time taken by the doctors to analyse the CT image of case is about 10 – 15 minutes but it can be quickly identified by CNN training CT image recognition network. AI need less than 20 seconds to identify a case on average, which greatly improves the diagnosis efficiency and overcome the burdens of doctors.

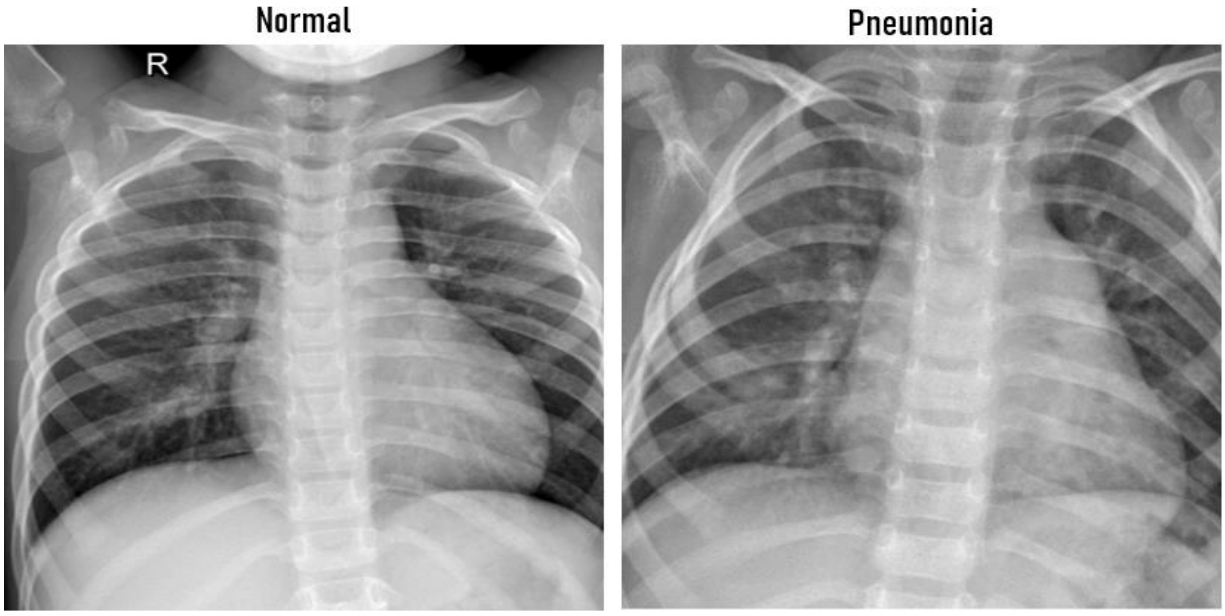
PROBLEM STATEMENT:

Pneumonia is a form of acute respiratory infection that affects the lungs. The lungs are made up of small sacs called alveoli, which fill with air when a healthy person breathes. When an individual has pneumonia, the alveoli are filled with pus and fluid, which makes breathing painful and limits oxygen intake. The medical establishment records that radiologist get it right on average only about 65 to 75% of the time and 5% of radiology interpretations contains clinically significant errors. The time taken by the doctors to analyse the CT image of case is about 10 – 15 minutes. So how to improve the prediction of pneumonia using x-ray in optimised way using Machine Learning. Chest X-ray imaging is the most frequently used method for diagnosing pneumonia. However, the examination of chest X-rays is a challenging task and is prone to subjective variability. Finally, something we are certain of is that ML is the next step of pneumonia detection.

Market / Customer/ Business Need Assessment:

Chest X-ray imaging is the most frequently used method for diagnosing pneumonia by Pulmonologist . The time taken by the doctors to analyse the CT image of case is about 10 – 15 minutes. AI need less than 20 seconds to identify a case on average, which greatly improves the diagnosis efficiency and overcome the burdens of doctors. Chest X-ray This helps your doctor diagnose pneumonia and determine the extent and location of the infection. But it can be quickly identified by CNN training CT image recognition network.So this depicts the use of optimisation treatment of pneumonia requirement for Pulmonologist.

It takes 15-20 minutes for direct detection of pneumonia for doctors.



ML finds it in 20 seconds .It's So quick.

Target Specifications and Characterization:

- To change traditional pneumonia detection process to faster and accurate process.
- Reducing frustration and death of patients due to delay in the detection process.
- Predetermined dataset of pneumonia patients and normal patients is taken and based on that prediction is performed.

Above, mentioned targets can be achieved by analyzing:

1. What the Doctor look for?
2. How are present detection processes are being performed ?
3. Problems faced by people suffering from pneumonia disease .
4. How to identify and provide treatment in initial stage accurately.
5. How efficiently are the Pulmonologist performing diagnosis process.

6. Analysing the needs of the patients suffering from pneumonia.

7. To help patient fight pneumonia at early stage.

8. To send results to the patient within minutes and prescribing the next step to be taken by the patient if he's been found of suffering from pneumonia.

Information searches:

Pneumonia is one of the diseases that people may encounter in any period of their lives. Approximately 18% of infectious diseases are caused by pneumonia. This disease may result in death in the following stages. In order to diagnose pneumonia as a medical condition, lung X-ray images are routinely examined by the field experts in the clinical practice.To develop a model that will help with the classification of chest X-ray medical images into normal(healthy) vs. abnormal(sick). To achieve this, seven existing state-of-the-art Machine Learning techniques and well-known Convolutional Neural Network models have been used to increase efficiency and accuracy. To achieve this, seven existing state-of-the-art Machine Learning techniques and well-known Convolutional Neural Network models have been used to increase efficiency and accuracy. In this study, we propose our Deep Learning for the classification task, which is trained with changed images, through multiple steps of pre-processing. Machine learning is a branch of artificial intelligence that employs a variety of statistical, probabilistic and optimization techniques that allows computers to "learn" from past examples and to detect hard-todiscern patterns from large, noisy or complex data sets. This capability is particularly well-suited to medical applications, especially those that depend on complex proteomic and genomic measurements. As a result, machine learning is frequently used in pneumonia diagnosis and detection.

Pneumonia dataset :

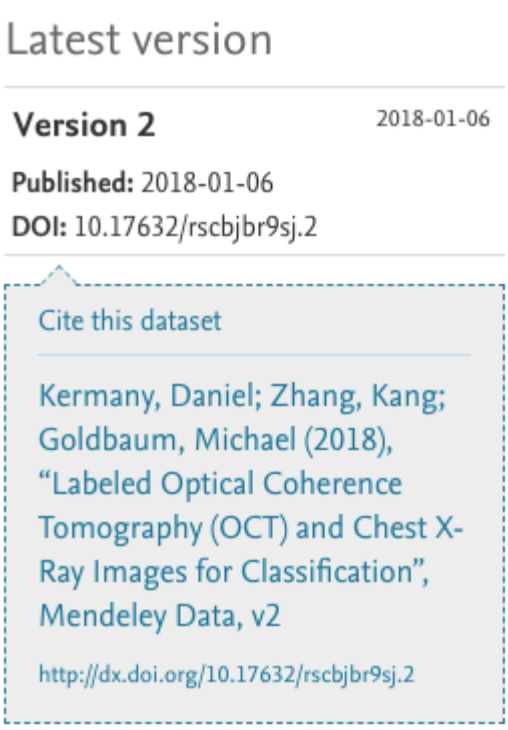
https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

- The dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal). There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal).
- Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care.
- For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system. In order to account for any grading errors, the evaluation set was also checked by a third expert.

Acknowledgements

Data: https://data.mendeley.com/datasets/rscbjbr9sj/2

Citation: http://www.cell.com/cell/fulltext/S0092-8674(18)30154-5



Inspiration

Automated methods to detect and classify human diseases from medical images.

Machine learning is the future of pneumonia prediction:

AI is set to change the medical industry in the coming decades — it wouldn't make sense for pulmonology to not be disrupted too. As datasets are getting larger and of higher quality, researchers are building increasingly accurate models. Machine Learning is the next step forward for us to overcome this hurdle and create a high accuracy pulmonology system.

Benchmarking alternate products:

Saliency methods, which produce heat maps that highlight the areas of the medical image that influence model prediction, are often presented to clinicians as an aid in diagnostic decision-making. However, rigorous investigation of the accuracy and reliability of these strategies is necessary before they are integrated into the clinical setting. In this work, we quantitatively evaluate seven saliency methods, including Grad-CAM, across multiple neural network architectures using two evaluation metrics. We establish the first human benchmark for chest X-ray segmentation in a multilabel classification set-up, and examine under what clinical conditions saliency maps might be more prone to failure in localizing important pathologies compared with a human expert benchmark. Fast and accurate diagnosis is critical for the triage and management of pneumonia, particularly in the current scenario of a COVID-19 pandemic, where this pathology is a major symptom of the infection. With the objective of providing tools for that purpose, this study assesses the potential of three textural image characterisation methods: radiomics, fractal dimension and the recently developed superpixel-based histon, as biomarkers to be used for training Artificial Intelligence (AI) models in order to detect pneumonia in chest X-ray images. Models generated from three different AI algorithms have been studied: **K-Nearest Neighbors, Support Vector Machine and Random Forest.**

Applicable Regulations (Government and Environmental) :

- Patents on ML algorithms developed
- Laws related to privacy for collecting data from users
- Protection/ownership regulations .
- Creating an e-mail service to mail the report to the patient and doctor.
- Being responsible by design.
- Ensuring open-source, academic and research community for an audit of Algorithms.
- Review of existing work authority regulations.

Applicable Constraints Expertise:

- Requires a lot of research to obtain universal dataset of cancer patients in-order to provide more sophisticated and accurate results.
- Establishing e-mail service in the product which have to send the report after the machine learning model is deployed in any server.
- Confidential health data to be obtained to train the model.
- Thorough understanding of dataset and verification of the results must be performed by the pathologist from the machine learning model to provide a great health prescription and service to the user.

Business Opportunity:

Pulmonologists are pretty good in diagnosing pneumonia . It takes more than 20 minutes weeks to identify pneumonia in an individual. To overcome this circumstance, our main objective is to use Machine Learning, which not only gives faster results but also demonstrates higher accuracy in the Pneumonia detection process.

Concept Generation:

The 4 stages of untreated lobar pneumonia are:

- Stage 1: Congestion
- Stage 2: Red hepatization
- Stage 3: Grey hepatization
- Stage 4: Resolution

Pneumonia is an infection in your lungs. It is a serious infection in which the air sacs (alveoli) in your lungs fill with pus, blood cells and other liquid. Almost all cases of pneumonia are caused by viral or bacterial infections. Lobar pneumonia is an acute bacterial infection of the lung and affects one or more sections (lobes) of your lungs.

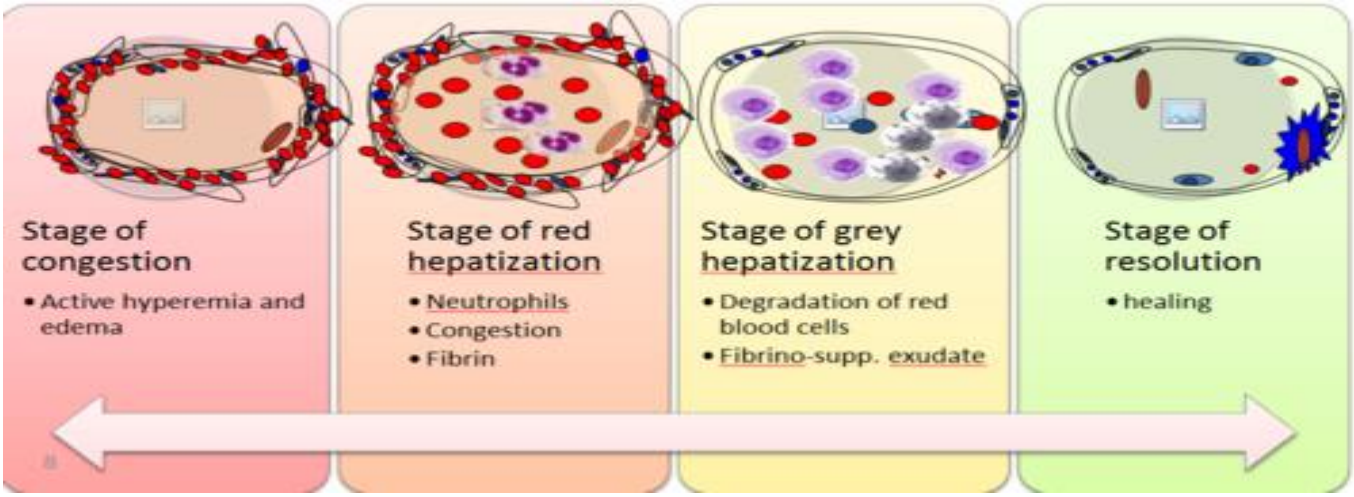
Stage 1 (congestion) occurs within 24 hours of infection. Many bacteria are present in the lungs but few white blood cells are available to fight the infection. The lungs may look red from increased blood flow and swelling of the lung tissue.

Stage 2 (red hepatization) occurs after 48 to 72 hours and lasts for about 2 to 4 days. The affected lung becomes more dry, granular and airless and resembles the consistency of liver. Red cells, white cells, bacteria and cellular debris can clog the lung airways.

Stage 3 (grey hepatization) occurs on day 4 to 6 and continues for 4 to 8 days. The lung looks grey or yellow in color but still has the consistency of liver. Fibrin, hemosiderin and red blood cells break down and lead to a more fluid-like exudate. Macrophages, a type of large white blood cell, start to form.

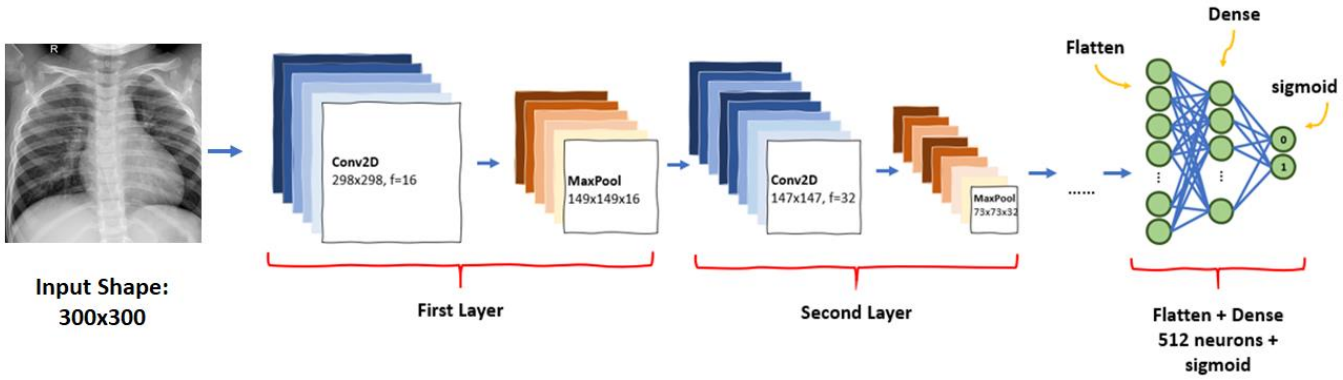
Stage 4 (resolution) is the final recovery stage and occurs during days 8 to 10. Fluids and breakdown products from cell destruction are reabsorbed. Macrophages (large white blood cells) are present and help to clear white blood cells (neutrophils) and leftover debris. You may cough up this debris. The airways and air sacs (alveoli) return to normal lung function. Any remaining lung swelling may lead to chronic lung disease (such as airway narrowing or pleural adhesions).

Pneumonia symptoms can include a cough with green, yellow or bloody phlegm or pus, chills, a fever, trouble breathing and shortness of breath, fatigue, chest pains and loss of appetite.

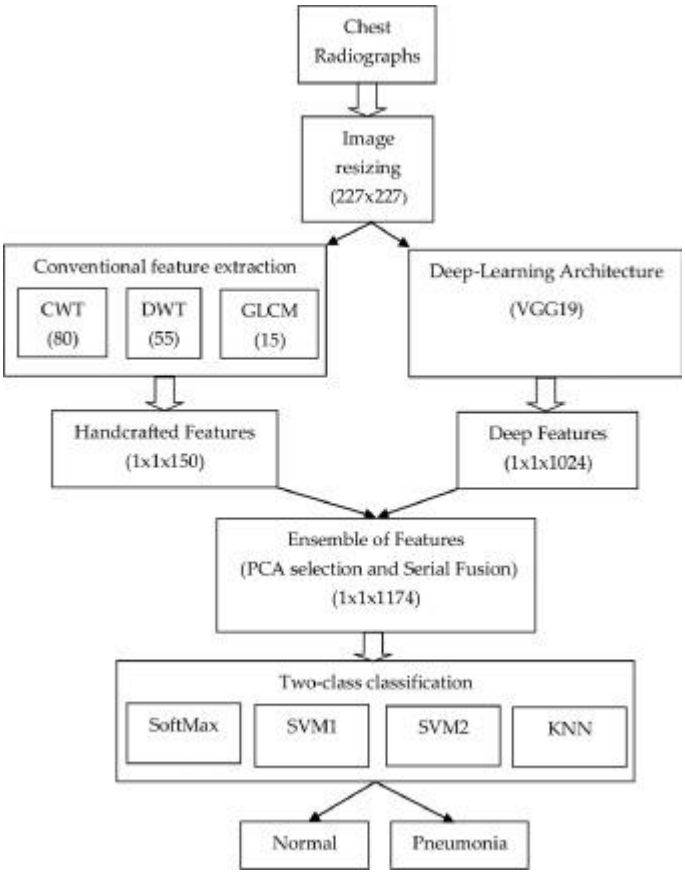


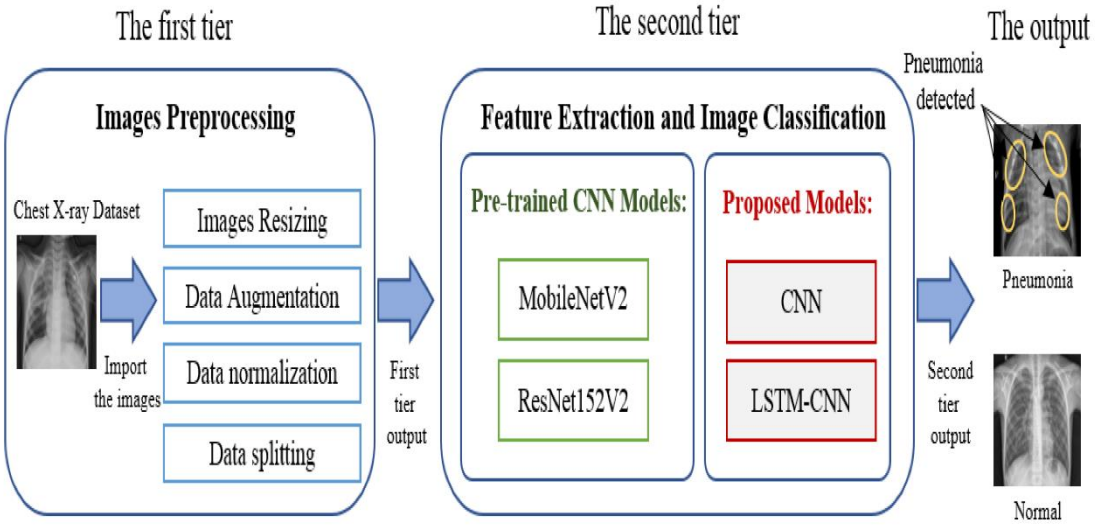
So in order to generate the model based on the problem stated above, we need to use Machine learning. Machine learning (ML) is the study of computer algorithms that improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so.

Pneumonia Detection using Convolutional Neural Network (CNN)



Final product prototype:



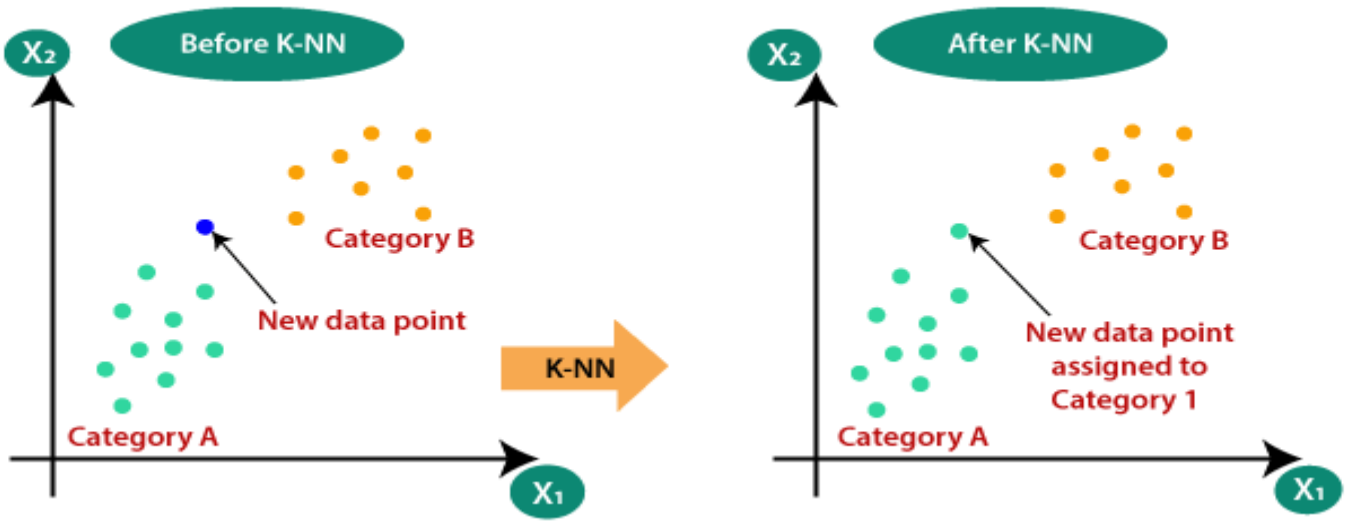


Algorithm:

Classification model contains different algorithms which can be used to classify chest x-ray based dataset.

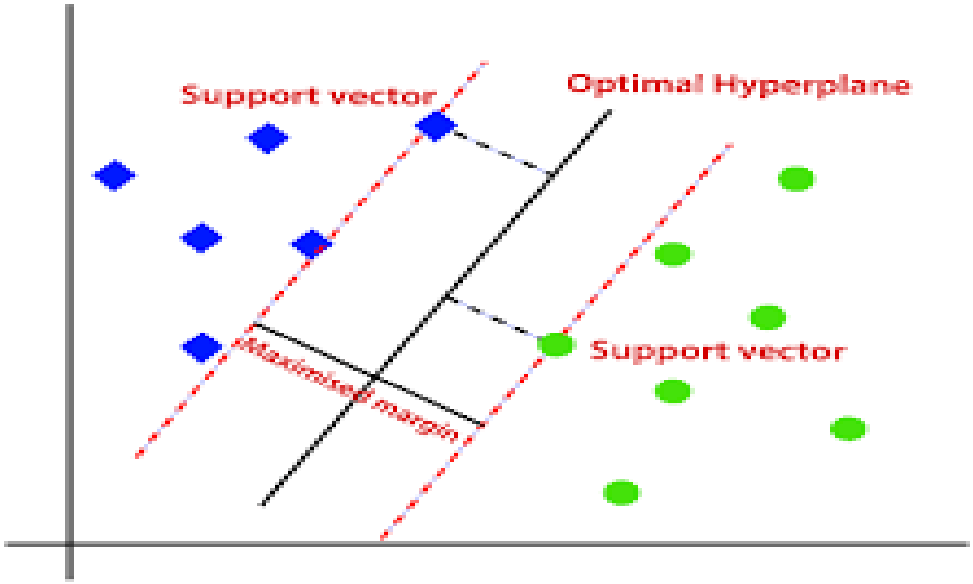
K-Nearest Neighbour

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
- It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.



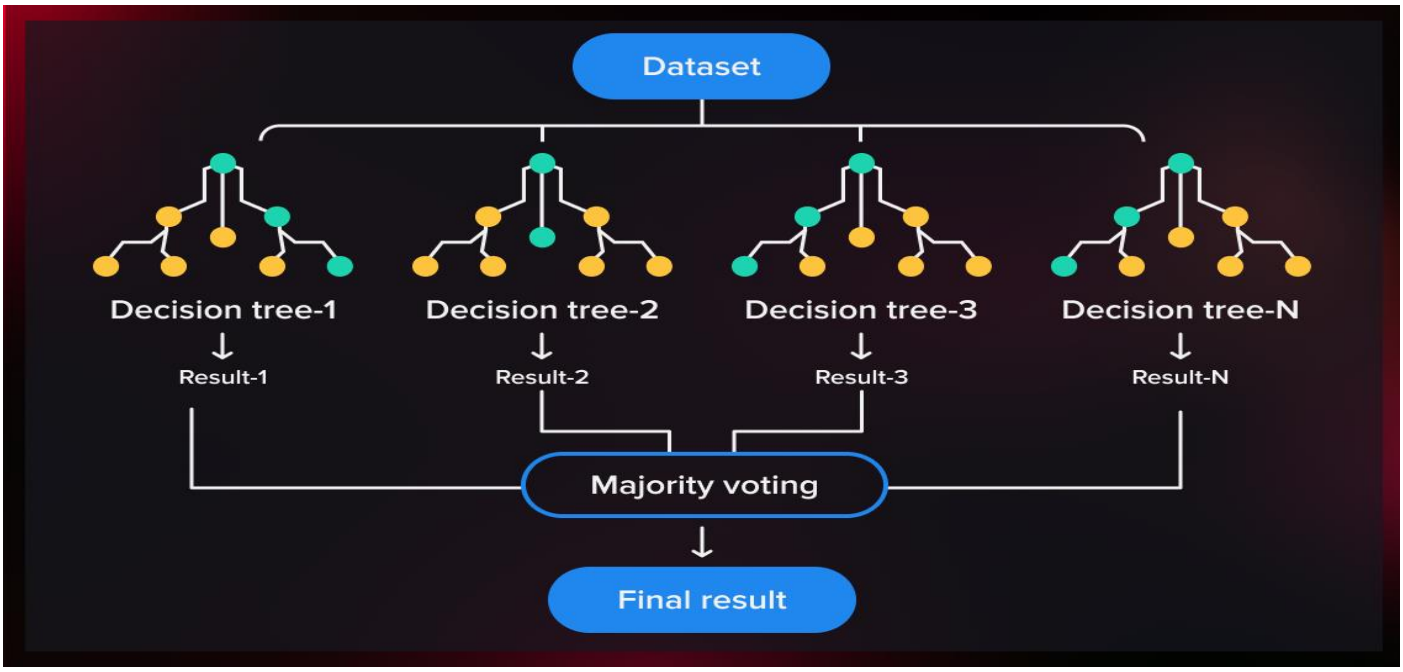
Support Vector Machine:

The goal of an SVM algorithm is to classify data by creating a boundary with the widest possible margin between itself and the data. The advantages of support vector machines are: Effective in high dimensional spaces. Still effective in cases where number of dimensions is greater than the number of samples. Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient Versatile: different Kernel functions can be specified for the decision function.



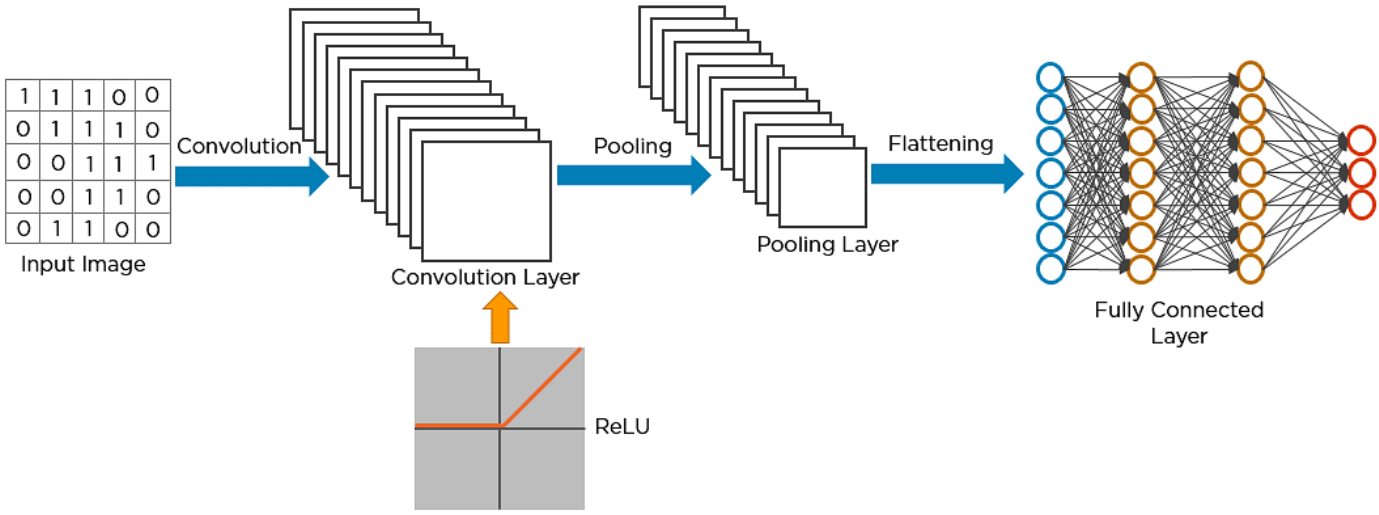
Random Forest:

The fundamental concept behind random forest is a simple but powerful one — the wisdom of crowds. In data science speak, the reason that the random forest model works so well is: A large number of 16 relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models. The low correlation between models is the key. Just like how investments with low correlations (like stocks and bonds) come together to form a portfolio that is greater than the sum of its parts, uncorrelated models can produce ensemble predictions that are more accurate than any of the individual predictions. The reason for this wonderful effect is that the trees protect each other from their individual errors. While some trees may be wrong, many other trees will be right, so as a group the trees are able to move in the correct direction.



convolutional neural network :

A convolutional neural network (CNN or convnet) is a subset of machine learning. It is one of the various types of artificial neural networks which are used for different applications and data types. A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data.There are other types of neural networks in deep learning, but for identifying and recognizing objects, CNNs are the network architecture of choice. This makes them highly suitable for computer vision (CV) tasks and for applications where object recognition is vital. The CNN is another type of neural network that can uncover key information in both time series and image data. For this reason, it is highly valuable for image-related tasks, such as image recognition, object classification and pattern recognition. To identify patterns within an image, a CNN leverages principles from linear algebra, such as matrix multiplication.



Team required to develop:

1. Machine learning engineering
2. Business analyst
3. Software developer
4. Data Researcher

Conclusion:

AI is set to change the medical industry in the coming decades — it wouldn't make sense for pathology to not be disrupted too. Currently, ML models are still in the testing and experimentation phase for pneumonia detection. As datasets are getting larger and of higher quality, researchers are building increasingly accurate models. While we might not see AI doing the job of a pathologist today, we can expect ML to replace our local pathologist in the coming decades, and it's pretty exciting! ML models still have a long way to go, most models still lack sufficient data and suffer from bias. Machine learning can train just as well as doctor prognosis, it doesn't require extra pay for prognosis. Manual pneumonia detection take long time to show the result ,while machine learning gives output in seconds .To save

people's life and allow doctor to fully concentrate in diagnosis.Yet, something we are certain of is that ML is the next step of pathology, and it will comfort the medical industry.

References:

1. www.google.com.
2. https://www.sciencedirect.com/science/article/pii/S1877050921008097?ref=pdf_download&fr=RR-2&rr=75ebe2115edf9a87
3. https://www.mdpi.com/2075-4418/10/9/649/htm
4. https://link.springer.com/article/10.1007/s11554-021-01086-y
5. https://ieeexplore.ieee.org/abstract/document/9208232