CS 445/545: Machine Learning, Fall 2021

Tuberculosis Detection Using Machine Learning Algorithms

README

Data source: Tuberculosis dataset
https://www.kaggle.com/raddar/chest-xrays-tuberculosis-from-india?select=jaypee_metadata.csv+–+Bing#

Language: Python

Platform: Jupyter Notebook
https://jupyter.org/install

Software Requirements: Mac OS or Windows 10 OS

Steps:

1) Download the dataset from Kaggle
2) Open Jupyter Notebook
3) Create a new notebook - python - name it as TB_Detection
4) In the notebook install the libraries and import them.
5) Get the data from the file that the dataset is stored in
6) Get the test data and print it out.
7) Print the format in which the data is present (image)
8) Trim the data and print the trimmed data.
9) Compress the data and print the data.
10) Apply the machine learning algorithm to get the prediction.
11) Compare the predictions of the algorithms and choose one that best suits the problem of Tuberculosis detection in the patients.
12) Print the results.

To do:

In this project we have used Local KNN, Weighted KNN and Gradient Boosting to predict the disease in patients and we have compared their accuracies and found the best for our problem. In the future, we can use other techniques like bagging, SVM technique for KNN and Boosting technique for better accuracy.