Medical Diagnosis Using Machine Learning

➢ **Paresh Patil**

Date:31-07-2022

1.Problem Statement

Decision making in case of medical diagnosis is a complicated process. Machine Learning techniques have been successfully utilized in a different application including to assist in medical diagnosis. It is very effortless and on time process for patients to analysed disease based on clinical and laboratory symptoms with appropriate data and give more efficient result for specific disease. Machine learning (ML) methods have been employed to assist clinicians in overcoming these limitations and in making informed and correct decisions in disease diagnosis. Many academic papers involving the use of machine learning for disease diagnosis have been increasingly getting published. Hence, to determine the use of ML to improve the diagnosis in varied medical disciplines. To carry out the review, one database is selected. Further, the database is classified as class, symptoms Then the selected database are analysed to show the impact of ML methods in improving the disease diagnosis.

2.Market/Customer/Business Need Assessment

Appropriate and effective treatment usually involves a thorough diagnosis These results will help in focusing on those areas which are neglected and also to determine various ways in which ML methods could be employed to achieve desirable results. These diagnostic errors could be minimized using techniques like fuzzy logic, and thus could improve healthcare services.

3.Target Specifications and Characterization

The main aim of this is to analyse the experiments in which ML approaches are used in relation to different medical symptoms their pattern and usefulness in the diagnosis of disease, through a systematic analysis.

4.External Search (Information Sources)

Using the Medical diagnosis Dataset for this project various types of datasets found for medical diagnosis. This dataset can be found on Kaggle

About Dataset

Link: https://www.kaggle.com/datasets/stevenmarrek/medical-diagnosis

```
In [1]: import pandas as pd
        import numpy as np
        import seaborn as sns
        import matplotlib.pyplot as plt
```

```
df.head()
```

	colored_sputum	cough	fever	headache	class
0	0.70	1.70	4.10	1.70	cold
1	0.38	3.22	5.20	3.21	cold
2	0.38	1.22	3.70	3.11	cold
3	0.36	3.06	3.19	2.64	cold
4	0.46	3.16	2.79	2.94	cold

```
df.describe()
```

	colored_sputum	cough	fever	headache
count	1631.000000	1629.000000	1629.000000	1630.000000
mean	4.034408	5.837716	8.200147	5.218669
std	2.629871	2.551540	2.077712	3.232034
min	0.000000	1.000000	1.100000	1.000000
25%	1.670000	3.520000	8.000000	2.492500
50%	2.950000	6.460000	8.740000	3.610000
75%	6.665000	7.780000	9.490000	8.510000
max	8.520000	10.640000	11.240000	11.930000

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1633 entries, 0 to 1632
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   colored_sputum  1631 non-null   float64
 1   cough           1629 non-null   float64
 2   fever           1629 non-null   float64
 3   headache        1630 non-null   float64
 4   class           1633 non-null   object 
dtypes: float64(4), object(1)
memory usage: 63.9+ KB
```

5.Benchmarking

```
df.shape
```

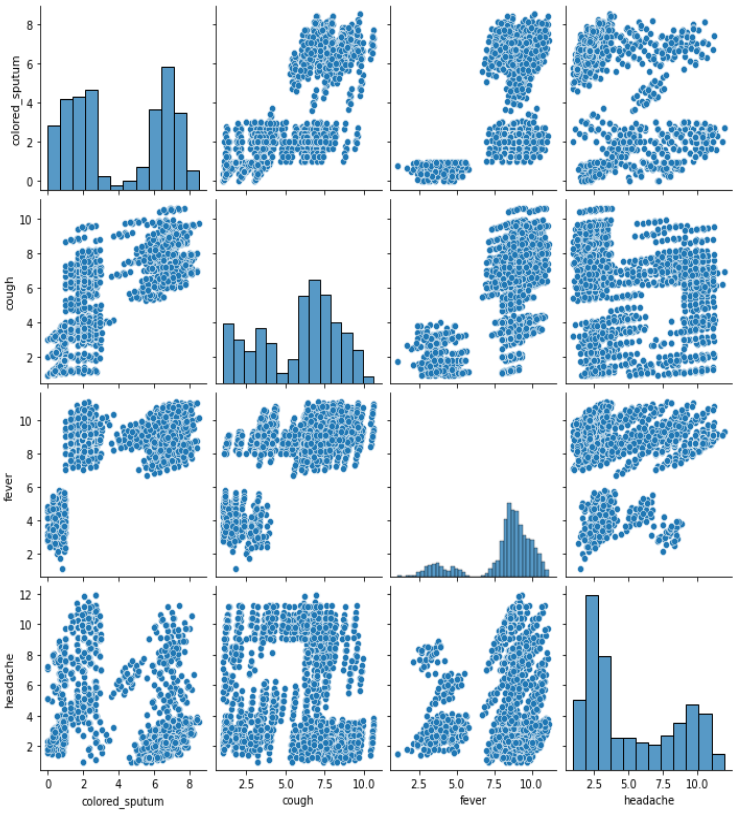
```
(1633, 5)
```

```
df.isnull().sum()
```

```
colored_sputum    2
cough             4
fever             4
headache          3
class             0
dtype: int64
```

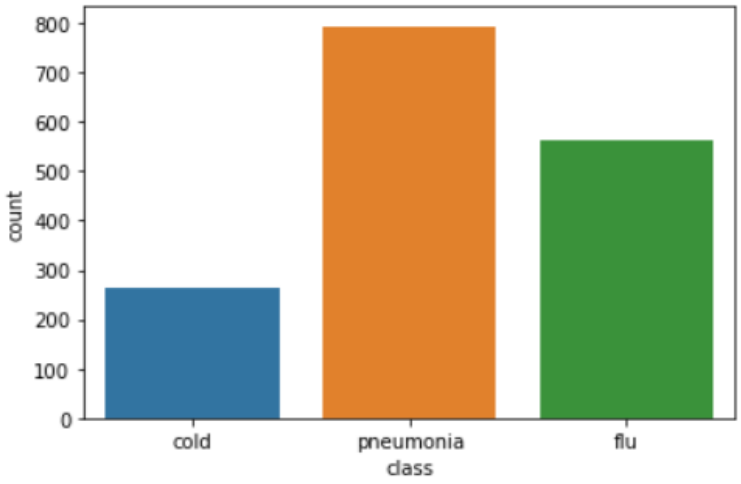
In [138]:
```
sns.pairplot(df)
```

Out[138]: <seaborn.axisgrid.PairGrid at 0x1e8a1138d90>

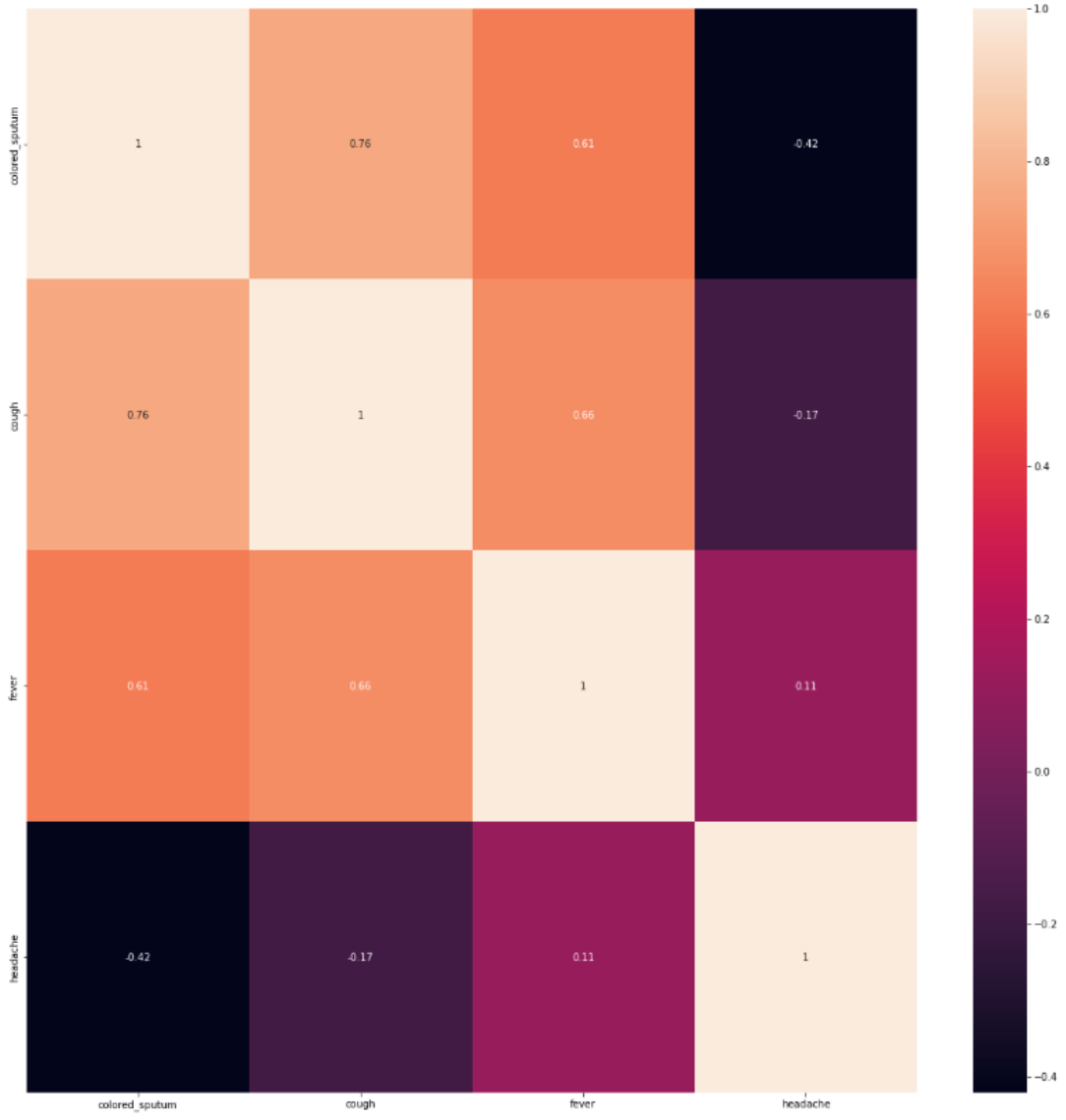


```
sns.countplot(x = 'class',data = df)
```

```
<AxesSubplot:xlabel='class', ylabel='count'>
```



In [139]:
```
corrmat = df.corr()
df1 = corrmat.index
plt.figure(figsize=(20,20))
#plot heat map
A = sns.heatmap(df[df1].corr(),annot = True,)
```



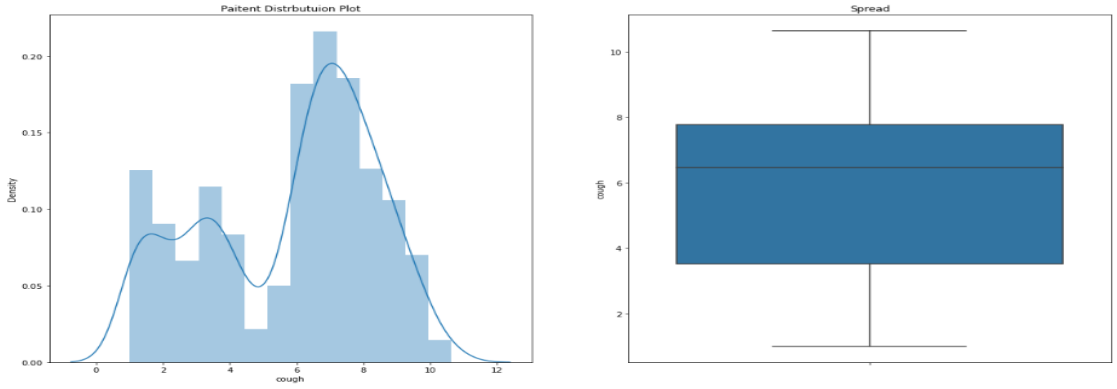
```
df.corr()
```

	colored_sputum	cough	fever	headache
colored_sputum	1.000000	0.757932	0.607555	-0.421251
cough	0.757932	1.000000	0.662710	-0.173477
fever	0.607555	0.662710	1.000000	0.112840
headache	-0.421251	-0.173477	0.112840	1.000000

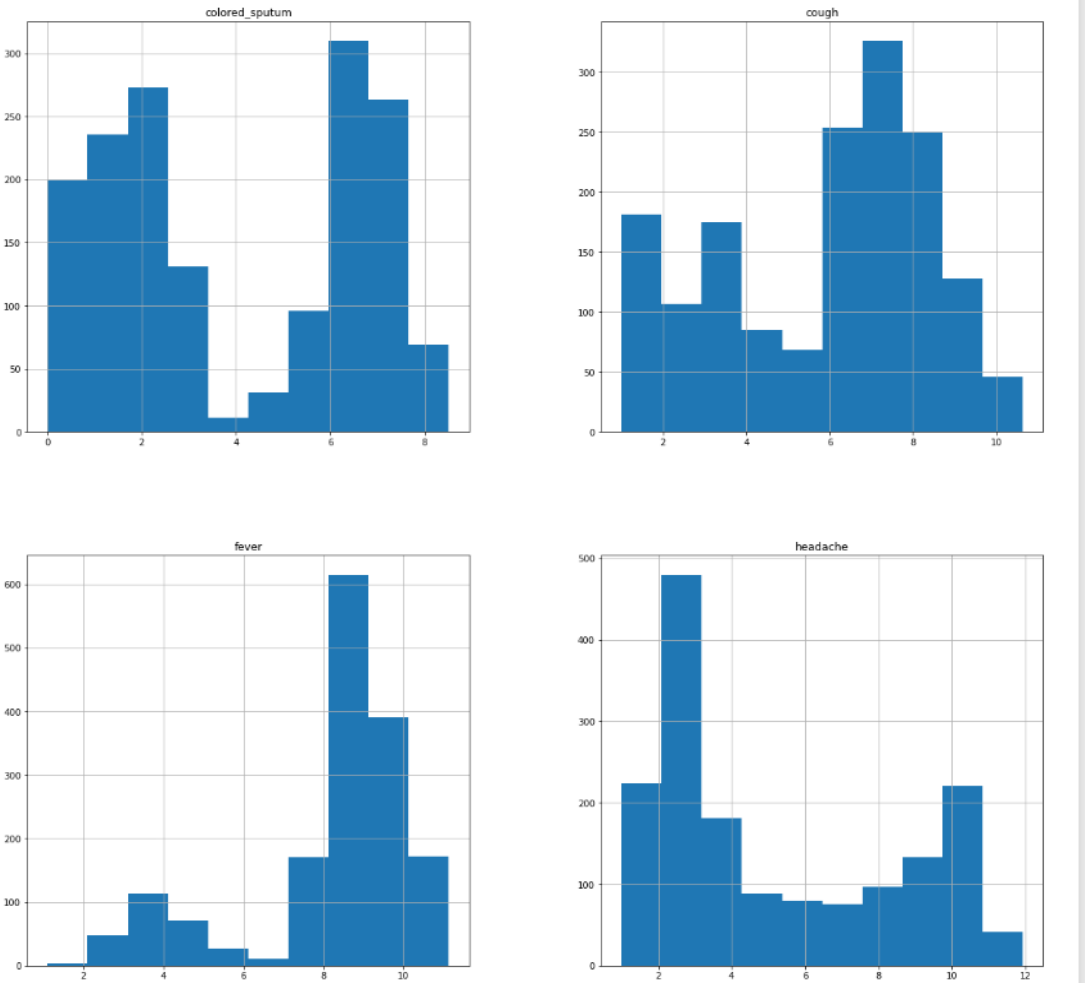
In [148]:
```
plt.figure(figsize = (20,10))
plt.subplot(121)
plt.title("Paitent Distrbutuion Plot")
sns.distplot(df.cough)

plt.subplot(122)
plt.title("Spread")
sns.boxplot(y = df.cough)
plt.show()
```

```
C:\Users\pares\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



In [146]:
```
#dist=df.hist(figsize=(15,15))
dist = df.hist(figsize=(20,20))
```



6.Applicable Regulations

- ✓ Legal requirements for the use of machine learning in medical devices(MDR and IVDR): - There are currently no laws or **harmonized standards** that specifically regulate the use of machine learning in medical devices. Obviously, these devices have to comply with existing regulatory requirements set out in the **MDR** and **IVDR**.
- ✓ Standards and best practices relevant to machine learning ("Artificial Intelligence in Healthcare" from the COICR): - The **COICR published the document *"Artificial Intelligence in Healthcare"*** in April 2019. It refers to existing requirements rather than providing new ones and recommends the development of standards.
- ✓ For more details refer link:-(https://www.johner-institute.com/articles/regulatory-affairs/and-more/regulatory-requirements-for-medical-devices-with-machine-learning/)

7. Applicable Patents

Title: -System and method for remote melanoma screening
Assignee: - Health Discovery Corporation
Publication date: - 2013-09-24

8. Applicable Constraints

If a new diagnostic tool is to be widely implemented in low-resource settings, the new method should be cheap, simple to use and not require laboratories or even stable power supply. While many new rapid diagnostic technologies may fulfil one or two of these criteria, few are able to fulfil them all. Unfortunately, much of the diagnostic research and development is performed in high-income countries, where these constraints are not always kept in mind. This way, many new and effective diagnostics like genomic analyses have been developed but are unfortunately not possible to use in low-resource settings.

9.Business Opportunities

Covid-19 has taught all of us to live a different lifestyle across the world. Many small business ideas are getting shut down now. One of the evergreen sectors is health care. There are enormous healthcare business opportunities in India. There is small business to moderate

business to large scale business which you can start in the healthcare segment.

10. Concept Generation

In current scenario, doctor is collecting all the record of the patient and based on that he will give medicines to patients. With this scenario, huge amount of time is wasted due to several reasons. By using machine learning classification algorithms, for any specific disease, we can improve the accuracy, speed, reliability and performance of the diagnostic on the current system. Machine learning is capable of offering automatic learning techniques to extract common patterns from realistic data and make sophisticate and accurate decisions.

11. Concept Development

The approach to build the movie recommendation engine consists of the following steps.

1. Perform Exploratory Data Analysis (EDA) on the data
2. Build the recommendation system
3. Get recommendations

12. Final Product Prototype

The distribution of various ML methods for diagnosis is analysed.

1. Perform Exploratory Data Analysis (EDA) on the data
2. Build the recommendation system
3. Select Suitable ML Algorithm
4. Distribute Data for Train and Test
5. Get recommendations.

13. Code Implementation

GitHub Link: -
https://github.com/PareshPatil1/Medical-Diagnosis/blob/main/Medical%20Diagnosis%20feeyn_labs.ipynb

14. Reference

- https://patents.google.com/patent/US20160350919A1/en
- https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/overview-device-regulation
- https://www.johner-institute.com/articles/regulatory-affairs/and-more/regulatory-requirements-for-medical-devices-with-machine-learning/
- https://www.kaggle.com/datasets/stevenmarrek/medical-diagnosis