BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Course Number : ME 366

Course Title : Electro-Mechanical System Design and Practice

Group : A18

Project Title:

Lungs Performance Analysis Using Spirometry

Submitted By:

Riasat Islam - 1710007

Sk Md Faysol Ahsan - 1710010

A N M Fuhadul Islam - 1710023

Department of Mechanical Engineering

Level-3 Term-1

Abstract:

About 65 million people suffer from various respiratory diseases, of which 3 million pass away each year, making respiratory diseases 3rd leading cause of death worldwide. But now, with the rising of COVID-19, this situation is deteriorating fast. One of the main dangers of respiratory diseases is that sometimes the patient collapse suddenly without showing any prior symptoms.

The objective of this project is to design an efficient yet affordable device that can predict the lung condition of the user within a brief time. If the condition is predicted to be bad, the user should visit the doctor for further consultancy. This device uses the method of spirometry to calculate the Forced vital capacity of the user. But this value does not indicate the lung condition as it depend on some other factors as well. So, the device will use a pretrained machine learning model to predict the lung condition of user. This model was trained by the method of linear regression over 1500 dataset and found to be accurate up to 90%. This device should be used as a precautionary step for regular lung performance monitoring.

Acknowledgement:

Acknowledgement: We would like to acknowledge the contribution of Asst. Prof. Dr. Md Aman Uddin and Asst. Prof. Musanna Galib of Mechanical Engineering department, BUET. Their constant feedback has helped us achieve a lot of insight and invigorated our motivation for the project. We'd also like to thank Rafia Rizwana Rahim for helping us with the ideas, and Sawradip Saha for aiding us in the development of machine learning model.

Table of Contents:

Abstract……………………………………………………………………………………..i

Acknowledgement…………………………………………………………………………ii

1.Introduction……………………………………………………………………………….1

2.Background……………………………………………………………………………….2

3.Literature Review…………………………………………………………………………3

4.Design and Methodology……………………………………………………………..10

5.Calculation………………………………………………………………………………14

6.Results and Discussion…………………………………………………………………17

7.Conclusion………………………………………………………………………………19

8.Reflexion on Learning…………………………………………………………………..19

9.Future Work……………………………………………………………………………..20

10.References……………………………………………………………………………...21

Appendix

1. Introduction:

Spirometry is a test to measure the volume of exhaled air. It is primarily used in medical settings to diagnose different pulmonary diseases such as asthma, chronic obstructive pulmonary disease. The goal of this project is to measure the “Forced Vital Capacity” of the lungs. “Forced Vital Capacity” or FVC is the maximum amount of air exhaled by a person. Spirometry gives us the value of this desired quantity. However, only getting the value of FVC is not enough to judge whether a person has healthy lungs or not. The reason being is that the value of FVC depends on many criteria such as height, weight, age, gender, race etc. So, to judge the healthiness of a person’s lungs, it is necessary to take into account of this factors.

In our project, the user of our device would be exhaling the air in an orifice tube where it will work as the mouthpiece. We will use fluid mechanics to analyze the flow and determine FVC. Then we will pass the value of FVC to a machine learning program to decide whether the user have healthy lungs or not.

2. Background

Spirometry could be a handy tool to survey lungs' performance continuously. The Center of Disease Control (CDC) has advised Covid patients to monitor their lungs performance using spirometers. During the pandemic, it will also be a helpful tool for patients who do not have readily access to expensive medical facilities in Bangladesh. In Bangladesh, medical fascilities is not enough for all people. And ICU's, specially outside of city corporations, is rare. Even inside city corporations hospital's capacity is not sufficient enough. So, for post-covid regular check up, we can't afford to monitor all the patients in hospital. So a low-cost way, which is affordable by all, is checking up lungs performance regularly by spirometry. According to WHO, "*About 1 of each 6 people who get covid-19 develop a serious illness and have difficulty breathing. Older people and those with underlying medical condition such as high blood pressure, heart conditions or diabetes, are more likely to develop serious illness.*" So elderly people and with other medical conditions, people should be more careful, as they are vulnerable to have serious illness. In this context, spirometry can provide high clinical information about patients respiratory conditions.

Lungs performance doesn't only depend on inhalation of exhalation air volume rate. With increasing age expiratory muscle strength, lung elasticity decreases and results a gradual decrease in FVC. Lung performances such as FVC, FEV are proportional to body size. With larger chest, lungs are larger, which have more and bigger alveoli, and that helps to have a great FVC. With higher height, chest size increases, which helps to have a greater FVC. Gender have a role in lungs normal condition also, as an adult female has lungs, 10-12% smaller than same height and weight's male lungs. Weight have good impact on respiratory system, until the BMI crosses healthy weight zone. When the BMI goes to overweight and then obese zone, then the excess body fat, compresses diaphragm, chest cavity and lungs. Then people who exercise regular, they have higher FVC rate than people who don't work physical works. Lung function have differences in different races, but the problem is there's no sufficient data to differentiate between all the races. So, generally it's said that, lung function varies among the main races of the world for whom sufficient data are available.

3. Literature Review:

3.1.Orifice Meter:

In this project, we are building an orifice flow meter. This will work as the mouthpiece. To analyze the exhaled air, we are going to use Bernoulli's equation.

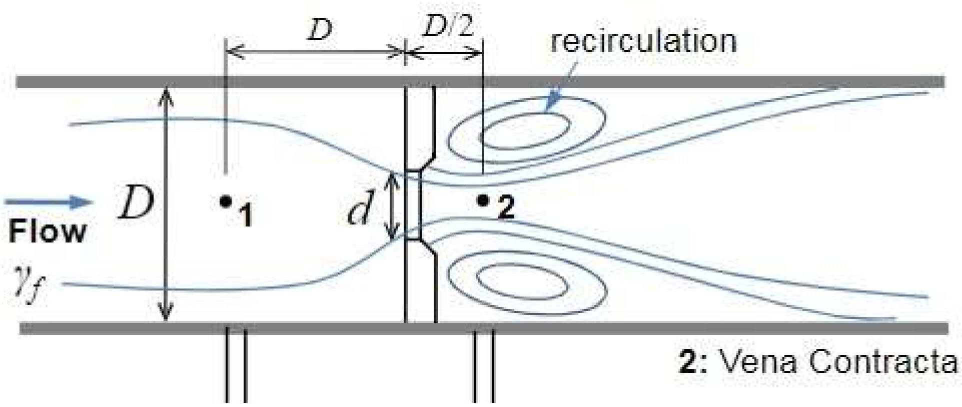


Figure1: Orifice Flow Meter

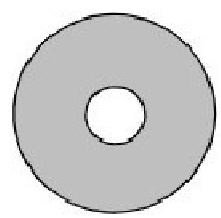


Figure2: Orifice Plate

We know that Bernoulli's equation is stated below-

$$\frac{p_1}{\gamma_f} + \frac{V_1^2}{2g} + z_1 = \frac{p_2}{\gamma_f} + \frac{V_2^2}{2g} + z_2$$

If we apply this equation for Orifice flow meter, the equation reduces to -

$$Q = \frac{A_o}{\sqrt{\left(1 - \frac{d^4}{D^4}\right)}} \sqrt{(2(p_1 - p_2)/\rho\,)}$$

Here,

Q = Volume flow rate

A_o = Area of the orifice

d = Diameter of the orifice

D = Diameter of the mouthpiece

P_1 = pressure in region 1

P_2 = pressure in region 2

ρ = density of the exhaled air

However, we should take into account the frictional losses the flow will have as it is being obstructed. Therefore, a discharge coefficient will be inducted into the formula of volume flow rate.

$$Q = \frac{C_d A_o}{\sqrt{\left(1 - \frac{d^4}{D^4}\right)}} \sqrt{(2(p_1 - p_2)/\rho\,)}$$

For ISO standard orifice flow meters. The value of C_d is around 0.6 to 0.66.

The equation is further reduced to-

$$Q = k\sqrt{(p_1 - p_2)}$$

Here, Volume flow rate is basically the pressure difference times a constant. The value of this constant is determined by calibration. This equation shows that, if we can find the pressure difference of the flow meter, we can easily calculate the volume flow rate.

To determine, FVC or the volume of the exhaled air, we would need to integrate the Volume flow rate over some time.

$$V = \int_0^t Q.\, dt$$

Now, we pass the value of V to out machine learning model to determine whether the FVC value is good to determine he/she has good pulmonary system.

3.2.Machine Learning:

3.2.1.Linear Regression:

In a linear regression model, the representation is a linear equation which give us a relation between some input features and an output. If the input is one feature and output is also one, then we can write the equation as : $y = mx + c$

Which we know is a straight line equation. But if the independent variable or input features is increased by one, then it becomes a equation of a plane:

$$z = ax+by+c$$

In general, we can say that it's a hyper plane equation when the total number of input and output is more than 2. In short, for machine learning we can say that a hyperplane is decision boundaries. So, for n number of input features we can write a hyperplane equation as

$$y = b_0 + m_1b_1 + m_2b_2 + ... m_nb_n$$

Where, y is the desired output. And m's are the input features or independent variables, and b's are their coefficients. We train a linear regression model with proper datasets.

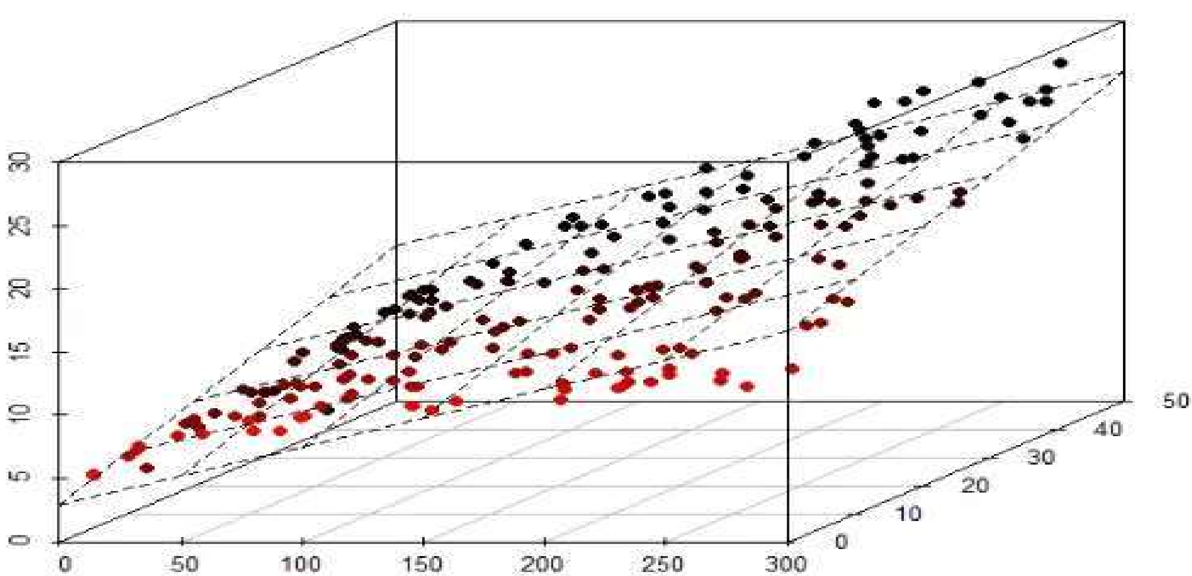


Figure3: Linear Regression in Hyperplane

3.2.2.Least Square Method:

In linear regression model, the approach used is least square method. It's a procedure of analyze the total sum of squares of differences between actual value and projected value.

If , $e_i = y_{i,actual} - y_{i,plotted}$

Then the least square method analyze of their squared sum.

$$S = \sum_{i=1}^{n} (ei)^2$$

And then find the smallest sum, for different coefficients of the plotted curve.

3.2.3.Gradient Descent:

In linear regression, with least square method, gradient descent is used to find the minimum error function. Gradient descent is an iterative method to analyze all weights with regard to small changes in error. It works by starting with random values for all coefficients, and then calculate the sum of squared errors for each pair of input and output values. Gradient is the measurement of how much the output of a function, with regard to error, changes with a little change in input. And gradient descent, starting with a random point of function, it moves in the negative direction of the gradient of the function.

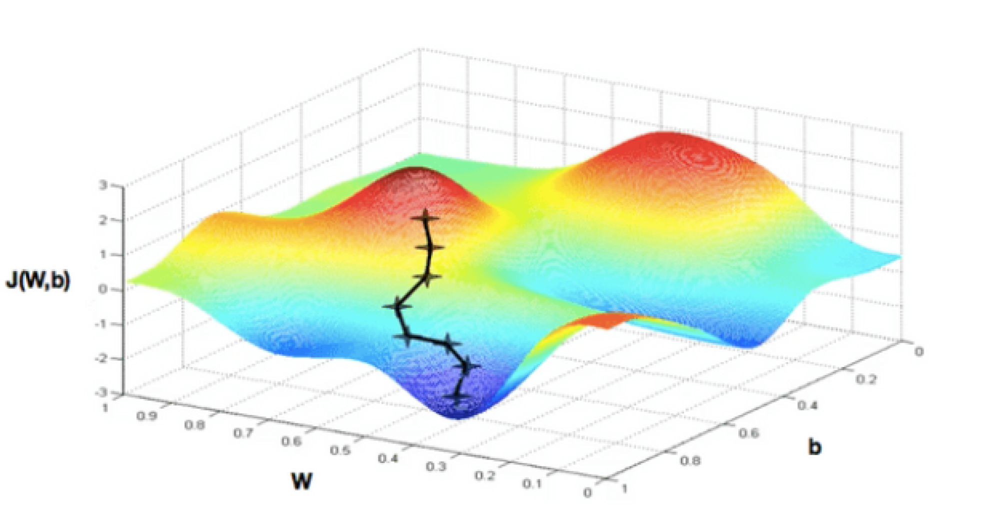


Figure4: Gradient Descent in Hyperplane

3.3.Components:

To make the project come into life, we have used numerous electrical components. Below is their name and their intended use:

3.3.1. Arduino Uno:

Arduino is an open-source electronics platform based on easy-to-use hardware and software. It is a single board microcontroller unit which contain the ATmega328 chip. In our project, the ports of the Arduino board are used as connecting wires to different electrical components. We also used it to power up our LCD and sensor. Most importantly, the Arduino is the brain of the whole process. It implements the necessary command and process the data to give out meaningful result.

3.3.2. Pressure Sensor:

The pressure sensor that we used in this project was “MPXV5050DP”.

Figure5: MPXV5050DP pressure sensor

It is a differential pressure sensor where there are duel ports. One port is connected before the orifice plate and another port is connected after the obstruction. The sensor basically determines the difference of pressure in these 2 region and give it as an analogue signal.

3.3.3. LCD Display:

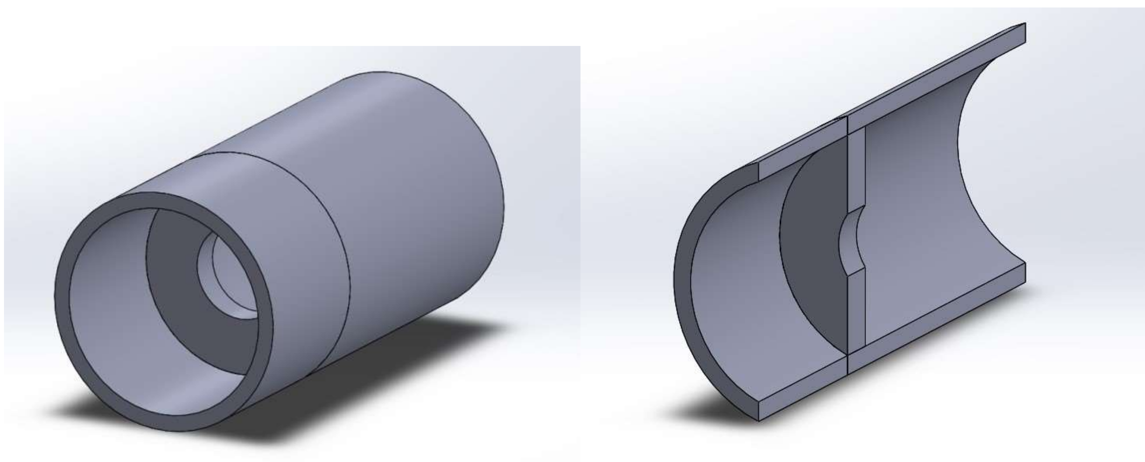
LCD display is primarily used a user interface, where it prompts the user to input the value of different features which are needed for the machine learning model. It also displays the healthyness of the lungs as output after the complete calculation. In this porject, we are using a 16x2 LCD display.

3.3.4. 3x4 Matrix Keypad:

We are also using a 3x4 matrix keypad here to input the values of different features that are needed to run the machine learning model. It works as an input device.

4. Design and Methodology:

For this project, we had to built a orifice flow meter which will work as mouthpiece. The diameter of the tube was decided to be 15mm and the diameter of the orifice hole was chosen to be 6mm. The reason to choose the diameter of this dimension is to make the mouthpiece accessible to both adults and children.



(a) (b)

Figure 6: CAD model of orifice flow meter tube (a) isometric view (b) sectional view

The pressure sensor is attached to the tube using two plastic tube which are attached to the two ports of the sensor. One tube goes before and another tube goes after the orifice plate. The pressure sensor is attached to the following circuit by an Arduino uno board-

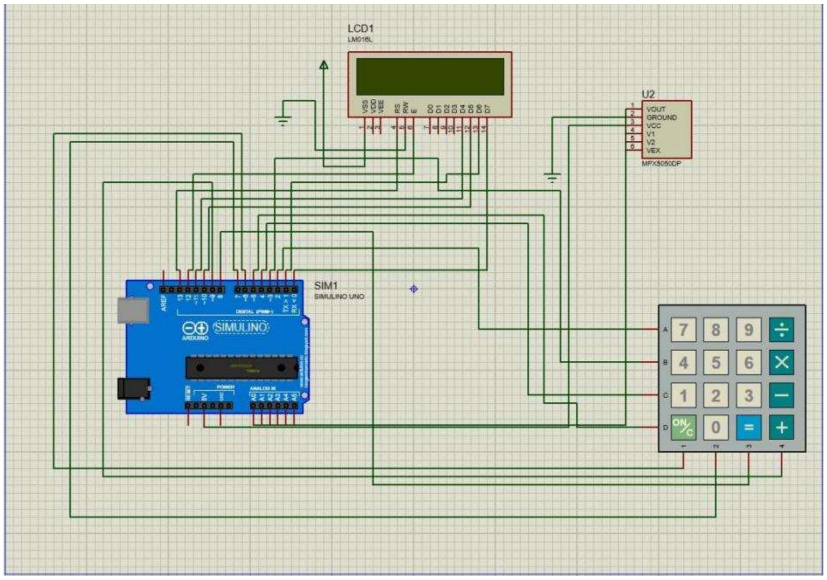


Figure7: Schematic Diagram of the Circuit

The LCD display and the keypad are also connected to the Arduino in the way that has been shown in the above figure. The Arduino is programmed to be run in the following way:

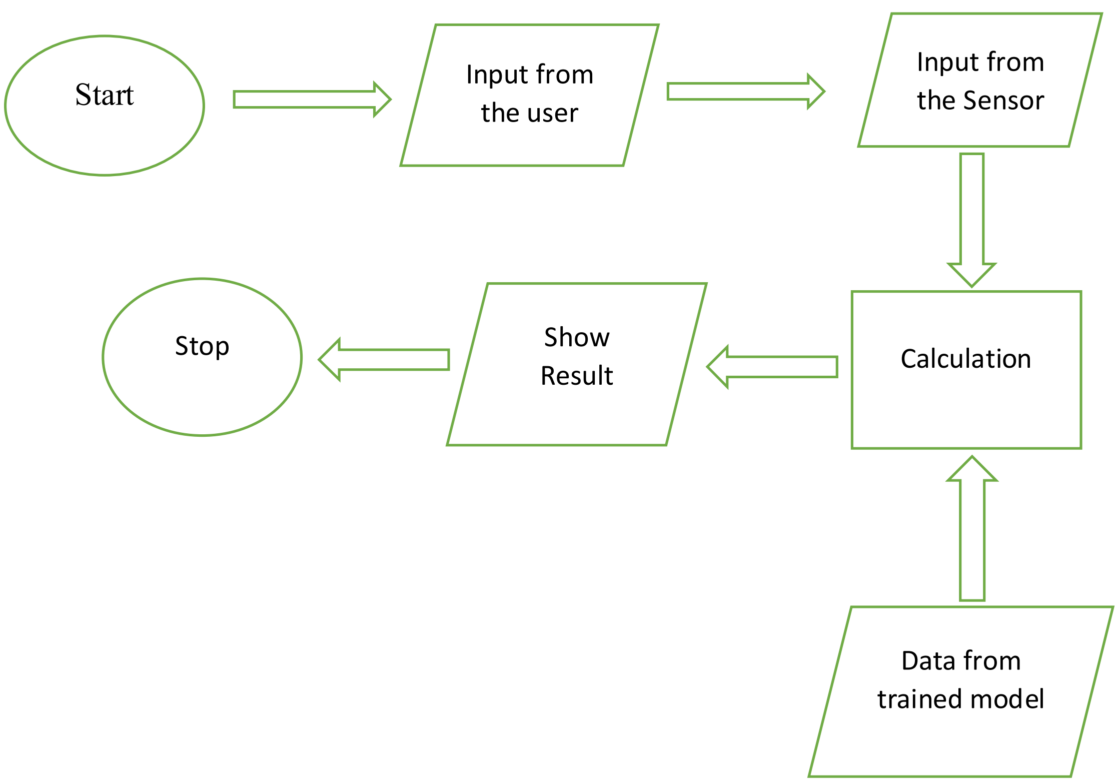


Figure8: System Flowchart

When the program is running, the lcd display prompts the user to entry 3 features, they are – "age", "gender" and "smoking history". Our pretrained machine learning model needs these features to analyze the Volume that the user will exhale. The features will be inputted with the help of the kay pad.

After inputting these values, the LCD display will ask the user to exhale air into the mouthpiece. The sensor will read the data and do necessary calculation to measure the Volume of the air. Then this value is processed by the machine learning model to give a final output of "Lungs' healthiness" in a percentage value.

4.1.Machine Learning Model Training:

4.1.1.Test Subjects:

To train a model, we have used spirometry data of total 1549 volunteers including 1224 male and 325 female. Whose age were between 18 to 88 years. We have also smoking status and FVC as input features in the data set.

4.1.2.Preparing Data:

To train a model, we used python programming language. After data input, we had to prepare the data for training our model, as in datasets there's so many errors and unwanted features. So, at first we had cleaned the data, made the data set error free, and stored only the necessary features for our model. Then there's some features(smoking status, gender) have text data, but for our calculation we had to convert them in numerical values, and at last,we divided the total data set into train data and test data in a ratio of 9:1. In python language, to manipulate data we have used python library numpy and pandas. Numpy and pandas library provide high performance operations, give way of working with multi-dimensional arrays and data manipulation.

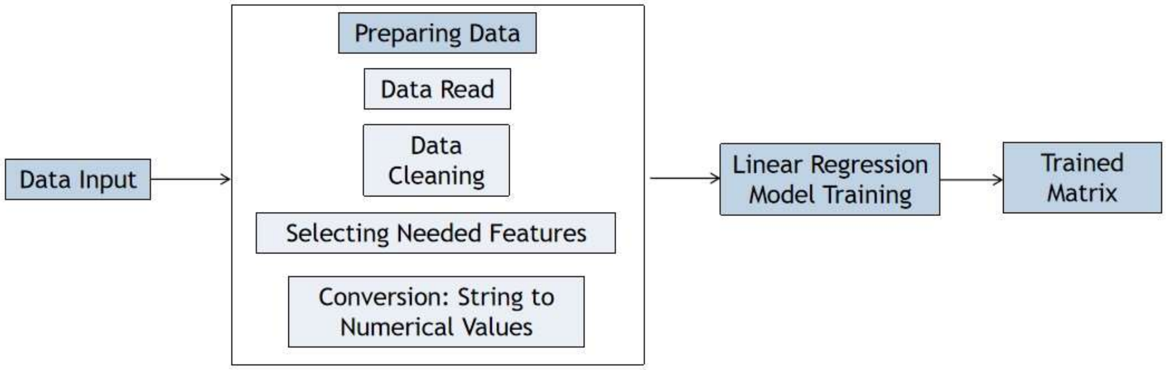


Figure 9: Schematic of ML Model Training

4.1.3.Model Training:

After preparing the data, we have trained a model by pythons linear regression model with test data. Then tested that model with test data. Test data and train data are totally independent of each other. We have discussed before about linear regression model. In python to use linear regression, we used scikit-learn library, which provides regression models and helps data manipulation. From a linear regression model, we have a weight(coefficient of the linear equation) matrix output. Which is integrated with Arduino Uno.

5. Calculations:

First part of calculating lungs' performance parameter is to calculate FVC (Forced Vital Capacity) from sensor data.

At first, the analog signal from the sensor has to be converted to digital signal. This is done with the equation

$$V = V_p \times 5/1024$$

Where V is the digital voltage in microcontroller and V_p is the analog voltage from sensor.

The second stage of this calculation involves converting this voltage to a physical quantity, differential pressure. This is done with the transfer equation of sensor

$$dP = (V/5-.04) \times 1000/0.018$$

Where dP is the differential Pressure in pascal unit.

Next, a modified form of the Bernoulli's equation converts this differential pressure to a differential volume flow rate dQ in the mL/s unit

$$dQ = (A_o \times ((2 \times P / \rho)^{0.5})/((1-B^4)^{0.5})) \times 10^6$$

Where A_o is the area of orifice opening in m^2, ρ is the density of air in kg/m^3, B is the ratio of diameter of orifice to the pipe.

Now, this dQ is the volume flowed over the time of only one reading. The sensor will be taking reading 100 times per second. To get the total volume during exhalation, this volume flow rate have to be summed over the time of full exhalation multiplied by the sampling time. Then the result will give the FVC of the user in mL unit.

To perform this summation, a loop is used in the microcontroller. This loop works as such:

Loop till dQ is greater than a trigger value:

$$FVC = FVC + dQ \times S$$

Here, S=0.01s, which is the sampling time. FVC is initialize to 0 at the start of the summation. The loop is controlled by a trigger value of dQ, which indicates the air flow rate in the orifice. If the value of dQ is lower than the trigger value, the loop terminates and uses the latest value of FVC for further calculations.

This trigger value depends upon the design of orifice meter. The usual lowest velocity of human exhalation is 2.2 m/s. then the trigger volume is expressed by the equation

$$\text{Trigger value} = 2.2 \times A_o$$

Here, A_o is the area of the orifice in m^2.

After these steps, the FVC of the user is available to use.

The next step is to calculate the lungs performance in percentage with a trained machine learning model.

The model used was trained by the method of linear regression. Then the weights of the input variables are taken from the model and stored in Arduino. These weights can be expressed by a matrix M as such:

$$M = \begin{bmatrix} -67.497798082503 & 0.0237200392 & 0.617981660 & 29.6510904 & 2.23838912 \end{bmatrix}$$

And, the data of the user can be expressed as a X matrix as:

X=

$$\begin{bmatrix} 1 & FVC & Age & Gender & \text{Smoking status} \end{bmatrix}$$

Then, the lungs' performance Y will be given by the equation :

$$Y=MX'$$

Here, as M and X both are 1×5 order matrix, multiplying M with X transposed will give a single numerical value which is the lungs performance parameter. But as only four features were used in the project, this same process was carried out by multiplying each weight with its corresponding variable. So, the equation used in this project is:

$$Y= m_0 +m_1 x_1 +m_2x_2 +m_3x_3 +m_4x_4$$

Here,

m_0, intercept of y axis = -67.497798082503

m_1, weight of user's FVC = 0.0237200392

m_2, weight of user's age = 0.617981660

m_3, weight of user's gender = 29.6510904

m_4, weight of user's smoking status =2.23838912

x1= FVC of user

x2=age of user

x3= gender of user

x4=smoking status of user

Thus, the lung performance parameter of user is calculated in percentage.

6. Results and discussions:

This project is designed to perform as a stand-alone device that can predict the lung condition of a user with provided data. After the lung performance parameter is known,
The user can compare his condition with this table and take precautionary measures.

Table 1: Table to judge lung condition from lung percentage parameter

Lung performance	Remarks
80% and above	Healthy lungs
Below 80%	Weak lungs, should visit a doctor

The machine learning model used in the project was trained in a computer with python rather than in Arduino UNO for three reasons:

1. Training any machine learning model in a 8-bit processor like the ATmega 328 which is being used in this project is very tedious and time consuming.
2. The micro controller used here is the Arduino UNO, which contains 2kB of memory. But the data set used in this project is 105kB which couldn't
   Be stored in Arduino's memory.
3. Python have wide range of inbuilt libraries and functions for machine learning processes which are readily available.

Also, another point that needs to be noticed is that to measure the flow rate, we have applied ideal conditions, not the formula for actual flow rate. The reason being is that, we were not able to calibrate our device against another standard flow meter to determine K due to the ongoing pandemic. That is why error are introduced in this calculation.

We used linear regression, as for our date, we have seen that the accuracy level doesn't change significantly in other machine learning models. So we used the easier one. From our trained model, we have outputs of about 85-90 accuracy. But we could not have train our model with sufficient data set, and could not add all the features on which lungs performance depends. With sufficient data and features we can have a model having an accuracy of about 99%.

Due to COVID-19 pandemic, we couldn't collect enough data from hospitals. So we had to depend only on online data. And that's not enough. And more ever in our collected data, we didn't have all the features affects lungs performance. Such as, weight, height, physical activities. So, if sufficient data are found, then the model can be improved more.

7.Conclusions:

The need for regular lung checkup for people suffering from respiratory diseases like asthma, pneumonia is of foremost importance. Moreover, as now a days COVID-19 cases are spiking, more and more people suffer from lung infection and post COVID respiratory implications. Unfortunately, in the developing countries the public health infrastructure is not capable to provide quality respiratory system checkup to the general population as the ratio of high-quality hospitals and skilled workforce with population is exceptionally low.Under this circumstance, the project emphasizes the fact that if these infections and implications are detected earlier, the patient can take required steps to avoid any ill-fated situation. This device can be manufactured in large numbers and is also cheap enough to be used as a personal checkup tool.This device uses an orifice meter to calculate the Forced vital capacity of the user with the help of Bernoulli's equation. Then it uses a machine learning algorithm to predict the lung performance parameter of the user depending on his age, FVC, gender and smoking habits with an accuracy of 85%-90%. This parameter indicates the lung condition of the user.

8.Reflection On Learning:

This project helped us to learn about

- To apply theoretical knowledge into real life scenarios, and how to improvise, adapt and overcome.
- Analyzing designs, and simulating of circuits in proteus and solidworks
- Machine learning algorithm and their implementations.
- Impact of our working habits on our respiratory system
- How to work in team environment and improve communication skills

9.Future Work

Implementing this project there's still some problems on which, research should be done in future.

The performance parameters of the components may change with time. So the spirometer should be calibrated after few days of uses. But calibration devices are very expensive and not available to all. Maybe this problem can be solved in a cheaper way by experimenting and making another calibration curve with time by the manufacturers.

In this COVID-19 situation, another problem is spirometry test often generate aerosols in the form of droplets due to patients coughing and therefore pose a considerable risk for the spread of infection to individuals and maybe the spirometer also. So some disinfectant mechanism and a mechanism for collecting the droplet should be researched.

Spirometry test needs co-operative patients, skilled staff, and repeated tests to ensure consistency. Some patients are not able to complete the test. So, research should be done to check if there's another way to complete the spirometry test such as using cough sound. As cough generation and forced exhalation have some similarities.

References:

1. Fluid Mechanics by Frank M. White

2. The Global Impact of Respiratory Disease – Second Edition Forum of International Respiratory Societies.

3. Roneel V. Sharan*, Udantha R. Abeyratne, Vinayak R. Swarnkar, Scott Claxton, Craig Hukins,and Paul Porter. - "predicting Spirometry Readings Using Cough Sound Features and Regression." https://doi.org/10.1088/1361-6579/aad948

4. Alejandro Talaminos Barroso,a Eduardo Márquez Martín,b Laura María Roa Romero,a,c Francisco Ortega Ruiz. "Factors Affecting Lung Function: A Review of the Literature" www.archbronconeumol.org

5. Vivek Agarwal, N.C.S. Ramachandran - "Design and development of a low-cost spirometer with an embedded web server."

Appendices:

- Abbreviation

  FVC = Forced Vital Capacity

  ML = Machine Learning

  LCD = Liquid Crystal Display

- Dataset

Datasets:

OSIC Pulmonary Fibrosis Progression (kaggle.com)

Link:

https://drive.google.com/file/d/1AiiPtdItXbrFgdq4oNI6EbKPA2lR3Guw/view?usp=sharing

Patient	Weeks	FVC	Percent	Age	Sex	SmokingStatus
ID00007637202177411956430	-4	2315	58.253649	79	Male	Ex-smoker
ID00007637202177411956430	5	2214	55.712129	79	Male	Ex-smoker
ID00007637202177411956430	7	2061	51.862104	79	Male	Ex-smoker
ID00007637202177411956430	9	2144	53.950679	79	Male	Ex-smoker
ID00007637202177411956430	11	2069	52.063412	79	Male	Ex-smoker
ID00007637202177411956430	17	2101	52.868646	79	Male	Ex-smoker
ID00007637202177411956430	29	2000	50.327126	79	Male	Ex-smoker
ID00007637202177411956430	41	2064	51.937594	79	Male	Ex-smoker
ID00007637202177411956430	57	2057	51.761449	79	Male	Ex-smoker
ID00009637202177434476278	8	3660	85.282878	69	Male	Ex-smoker
ID00009637202177434476278	9	3610	84.117812	69	Male	Ex-smoker
ID00009637202177434476278	11	3895	90.758691	69	Male	Ex-smoker
ID00009637202177434476278	13	3759	87.58971	69	Male	Ex-smoker
ID00009637202177434476278	15	3639	84.79355	69	Male	Ex-smoker
ID00009637202177434476278	22	3578	83.372169	69	Male	Ex-smoker

Desired Features	
Unwanted Columns	

- Machine Learning Code:

  https://www.kaggle.com/faysolahsan/spirometer-ml

- Physical Demonstration:

Video Link:

https://buetedu-
my.sharepoint.com/:v:/g/personal/1710023_me_buet_ac_bd/EWinwMPboH9Dl1zgDj6Ot5QBs8
nqDN0oW2PLSoA6PbNQJA?e=Z9Q4qc

a b

Figure: a. inside view(circuit connctions), b.user Interface

- Arduino Code, CAD files, Circuit Simulation

  Link:

  https://buetedu-
  my.sharepoint.com/:f:/g/personal/1710023_me_buet_ac_bd/EptwLVlksLdDnQZhf2z3YDk
  BtjdV_wilTVFiqyhsVwu7DQ?e=Dvg9LM