IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

REGRESSION MODEL TO DETECT, PREDICT OXYGEN AND HOSPITALIZATION REQUIREMENTS OF COVID-19 PATIENTS

[1]Monika Sree Saankhya Kathula, [2]Pallavi Priya Kadiyam, [3]Vaishnavi Kadapathri, [4]Dr.Sunil Bhutada,

[123]B.Tech Student, [4]Professor

[1]Information Technology,

[1]Sreenidhi Institute of Science and Technology, Hyderabad, India

Abstract: COVID-19 pandemic is the deadliest crisis the entire world is going through. There are many active cases and deaths that have been brought about by the infection. The main symptoms of the disease include fever, dry cough, sore throat, tiredness and difficulty in breathing. To prevent the spread of disease, infected people should maintain hygiene and to be kept in isolation. Masks, sanitiser, and physical distance are to be maintained to control the spread of disease. This disease mainly affects the respiratory system thereby affecting the oxygen intake which is essential for human survival. In the present paperwork, we estimated the probability of a patient being affected with COVID-19 and estimated whether they need oxygen or should be hospitalized or if home quarantine is enough based on the given patient's details.

Index Terms **- Logistic Regression, Django, Python, COVID-19, Symptoms, SPO2, Hospitalization.**

I. INTRODUCTION

COVID-19 spreads from one person to a different through the air. Many countries have implemented lockdowns. The world has faced many economic issues. Several apps are created by the Government for aiding the people in fighting against the virus. Information like COVID cases, high-risk areas and information associated with COVID patients can be tracked using these apps. But a number of these haven't been effective due to privacy issues and lack of enough data to give accurate results.

In this project, we developed a website to help patients and the health care personnel to diagnose COVID-19 better along with minimizing the contact between patients and health care personnel while communicating about their symptoms, test results and personal details which are required for diagnosis.

According to ICMR, the total number of tests conducted in India as of 1st June, 2021 is 35,00,57,330 and total number of positive cases are 2,81,75,044 that is on an average only 8.04% of tests are conducting are showing positive results. This is a problem because even when the number of people affected by COVID were increasing rapidly the number of tests were not increasing at the same rate.

In this project we are estimating the probability of a person showing a positive COVID result so that it will help the patients to detect whether the symptoms they are exhibiting correspond to COVID or not and it also helps the respective authorities to provide tests for the most deserving people who have high probability of showing positive result, thereby decreasing the wasted COVID tests, which are already scarce.

Based on the information given by the patient on our website we are determining the probability of a patient affected by COVID-19 by applying machine learning techniques like logistic regression. We are also predicting the oxygen or hospitalization requirements based on the data given by the patient.

This data and results can be printed, saved or scanned from the QRcode we are generating and can be shared to require people digitally without any physical contact.

II. LITERARY SURVEY:

●In June 2020,Sonia Shah DO, Kaushal Majmudar DO, Amy Stein PhD, Nita Gupta MD, Spencer Suppes DO, Marina Karamanis DO, Joseph Capannari DO, Sanjay Sethi MD, Christine Patte DO proposed a work on "Novel Use of Home Pulse Oximetry Monitoring in COVID-19 Patients Discharged From the Emergency Department Identifies Need for Hospitalization".They predominantly centered around assessment of patient-revealed SPO2 utilizing pulseoximeter as a home observing apparatus for patients with first non-serious COVID-19 to distinguish need for hospitalization.

●In May 2020,Yong Li, Kun Zhao, Hongcheng Wei, Wensen Chen, Wei Wang, Ling Jia, Qiongfang Liu, Jinpeng Zhang, Tao Shan, Zhihang Peng, Yun Liu, Xiaoxiang Yan proposed work on "Unique connection between D-dimer and COVID-19 severity".They have examined the powerful connection between D- dimer level and the movement of COVID-19.It is a lot of filed that strange D- dimer is helpful in showing significant venous circulatory trouble in cardiovascular illnesses. By and large when all is said in done, dynamical changes of D-dimer level are quite related with the conjecture of COVID-19.

- In July 2020,Thomas Struyf,Jonathan J Deeks,Jacqueline Dinnes,Yemisi Takwoingi,Clare Davenport proposed a work on "Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease". To evaluate the analytic precision of signs and side effects to decide whether an individual introduced in essential consideration or to medical clinic outpatient settings, like the crisis office or committed COVID-19 centers, has COVID-19 infection or COVID-19 pneumonia. It concluded that based on currently available data, neither absence or presence of signs or symptoms are accurate enough to rule in or rule out disease.

- InAug 2020,Gökçen Budak, Oğuz Karabay, Ertuğrul Güçlü, Hüseyin Doğuş Okan, Aslı Vatan proposed a work on "Main symptoms in patients presenting in the COVID-19 period". Tests were acquired from the patients and PCR tests were played out; the patients were gathered as COVID-19 positive and COVID-19 negative; these two sets were addressed for 15 questions regarding symptoms and the outcomes are analysed. They reasoned that taste and smell issues were huge markers in COVID-19 defilement.

III. DATASETS

For estimating the oxygen or hospitalization requirements of the patients we have collected data about the patients tests details such as comorbidities, respiratory problems, Number of breaths per minute, SPO2 level, Pulse rate per minute, RT PCR test results(in cycles), HR-CT(CO-RADS levels 1-6), HR-CT severity index(out of 25), D-dimer(in µg/ml),CRP(c-reactive protein test)(in mg/L), ESR(erythrocyte sedimentation rate)(in mm/hr), RBC(Red Blood Cells), WBC(White Blood Cells) and Blood Platelets, Vaccination dose-1, Vaccination dose-2, oxygen requirements, hospitalization, home quarantine.

	noofdayssi	Age	comorbodi	respiratory	noofbreath	spo2	pulse	rtpcr	hrct	hrct-severi	d-dimer	crp	esr	rbc	wbc	platelets	dose1	dose2	oxygenreq	hospitaliz	homequare
2	2	35	0	0	19	96	82	35	2	9	0.18	0.5	17	4.8	6612	3.25	1	1	0	0	0
3	7	20	0	0	17	97	91	28	4	15	0.23	0.7	22	5.2	8841	2.5	1	0	0	0	1
4	4	45	0	1	14	96.32	98	9	5	21	1.12	1.42	54	4.6	5214	1.2	1	0	0	1	0
5	14	16	0	1	27	92	99	8	4	19	1.26	0.8	28	5.2	4125	1.25	0	0	1	1	0
6	12	55	1	0	29	91.57	96	7	5	23	1.12	1.42	40	4.8	6254	1.7	0	0	1	1	0
7	6	38	0	1	28	89	102	6	5	14	0.54	0.8	24	4.5	4125	1.2	0	0	1	1	0
8	9	32	0	1	35	89	91	9	5	23	1.05	1.18	30	3.5	5621	1.15	0	0	1	1	0
9	8	62	1	1	22	95	101	6	4	24	1.1	1.12	52	5.1	8112	2.1	1	0	0	1	0
10	4	22	0	0	18	89.62	98	21	4	14	0.84	0.8	19	4.9	5628	3.6	1	0	1	0	1
11	11	50	0	0	31	81	91	9	4	18	0.75	0.72	21	5.4	5432	3.4	1	0	0	1	0
12	6	26	0	1	29	89	73	15	6	25	1.55	0.91	26	3.2	3990	1.2	0	0	1	1	0
13	8	18	0	1	12	94.68	62	20	4	10	0.26	0.7	21	4.6	4110	4.02	1	1	0	0	1
14	5	33	1	0	25	96	88	7	3	9	0.31	0.7	21	4.5	5263	3.9	0	0	1	1	0
15	2	47	0	1	33	95.86	98	8	4	11	0.33	0.9	24	4.7	5412	3.8	0	0	1	1	0
16	13	52	1	1	11	96.14	92	7	4	12	1.52	0.7	26	5.1	7021	3.1	0	0	1	1	0
17	5	52	0	0	17	61	74	16	4	18	0.94	0.7	23	5.2	6013	2.73	1	1	0	0	1
18	3	23	0	1	27	67.83	91	8	6	24	1.52	0.98	25	3.2	3842	1.24	0	0	1	1	0
19	4	81	0	0	19	76.54	69	15	4	18	0.69	0.7	23	5.2	6013	2.73	1	1	0	0	1
20	9	46	1	1	32	62.3	82	8	6	24	1.21	0.81	25	3.7	4567	1.2	0	0	1	1	0
21	14	62	0	0	14	66	65	22	4	18	0.76	0.81	26	5.1	7058	2.69	1	0	0	0	1
22	2	77	1	1	24	97.26	63	15	4	12	0.36	0.9	22	5.2	6300	3.6	1	0	0	1	0

Figure 1)COVID tests dataset

For estimating the COVID-19 probability using symptoms we have used the "Patient Medical Data for Novel Coronavirus COVID-19" dataset from the Wolfram repository. It contains patient record including age, sex, location, date of onset, symptoms, travel history, chronic diseases, and date of discharge or death. We have manually cleaned the dataset and assigned binary values for each symptom such that 1 indicates the patient is exhibiting the symptom and 0 otherwise.

	AGE	GENDER	FEVER	COUGH	FATIGUE	PAINS	NASAL_CO	SHORTNESS	RUNNY_NC	SORE THRO	DIARRHEA	CHILLS	HEADACHE	VOMITING	LIVES_IN_A	COVID_OUTPUT
2	44	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1
3	45	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1
4	66	1	1	1	0	0	0	0	0	0	0	0	1	0	1	1
5	45	1	1	1	0	0	0	0	0	0	0	0	1	0	1	1
6	33	1	1	1	0	0	0	0	0	0	0	0	1	0	1	1
7	32	1	1	0	0	0	0	0	0	0	0	0	1	0	1	1
8	38	1	1	1	0	0	0	0	0	1	0	0	1	0	1	1
9	56	1	1	0	0	0	0	0	0	0	0	0	1	0	1	1
10	61	0	1	0	0	0	0	0	0	1	0	1	1	0	1	1
11	63	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0
12	64	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
13	50	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0
14	45	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
15	52	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
16	65	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
17	65	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
18	82	0	1	0	0	1	1	0	0	0	0	0	0	0	1	1
19	72	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1
20	39	0	1	0	0	0	0	0	1	0	0	0	0	0	1	1
21	27	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
22	62	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1
23	66	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

Figure 2)COVID symptoms dataset

IV. ARCHITECTURE

The front end of this project is a website that is created using basic Web Technologies like HTML, Materialize CSS, BootStrap, JavaScript. A high-level Python web framework Django helped us to run machine learning models which were developed in Python in the backend of our project. The user fills in all the required details on the website and we process them by applying Logistic Regression model which gives us the required results. We are dividing the dataset into 2 parts. 80% of the data is used for training and the other 20% of the data is used for testing purposes.The results generated are displayed to the user so that he can save/print or share the results and his details.

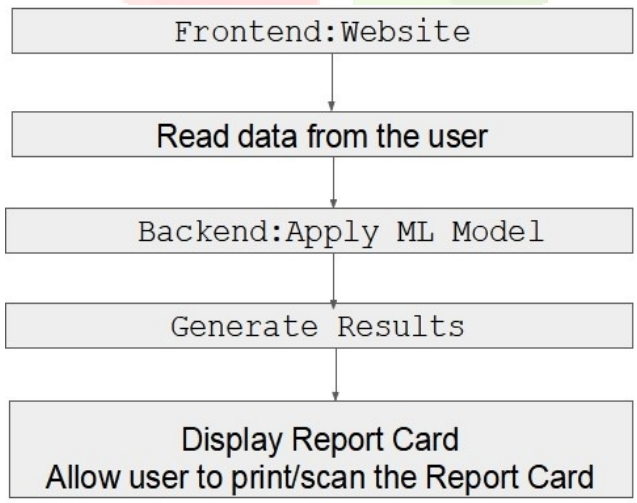

Figure 3)Architecture Diagram

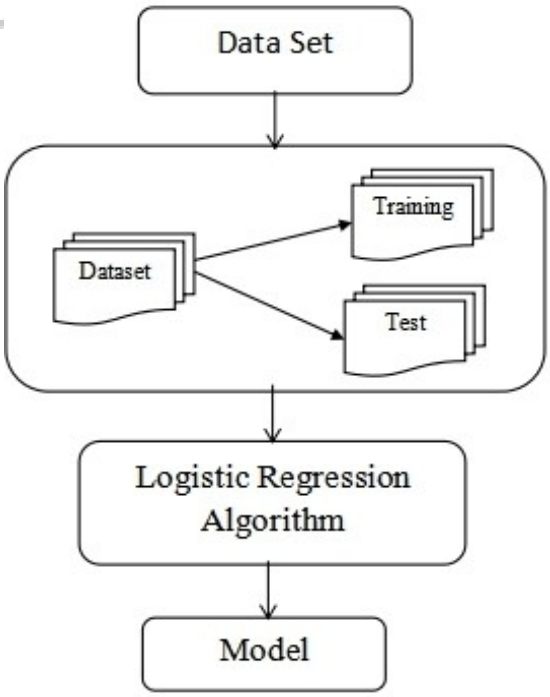

Figure 4)Architecture diagram of Machine Learning part

V. LOGISTIC REGRESSION

Logistic regression is a statistical model which uses logistic function to model a binary dependent variable based on multiple binary, categorical or continuous explanatory variables. Logistic regression is very much like linear regression but the only difference is that linear regression gives continuous values as output whereas logistic regression gives a binary output. There are three types of logistic regression: binomial, multinomial and ordinal. In this project, we used binomial because there are only two possible outputs either 1 or 0. We are using predict_proba() function available in scikit learn to calculate the probability of the outcome being 1.

In other terms logistic function is also called sigmoid function. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits. Here S(x) is the Sigmoid/Logistic function, e is the Euler's Constant and x is the value we are applying the function on.

$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^{x}}{e^{x} + 1}$$

Figure 5)Sigmoid/Logistic function

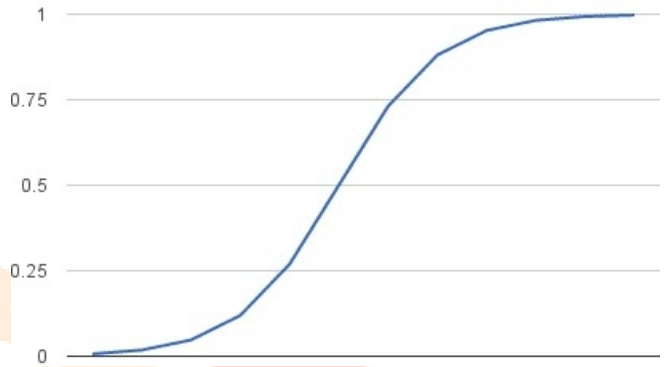

Figure 6)Logistic function graph

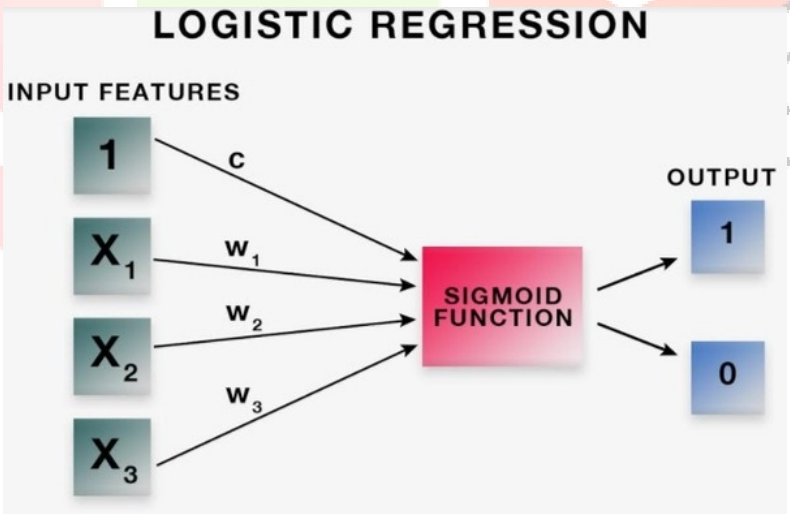

Figure 7)Logistic regression

VI. RESULTS

Webpage for COVID-19 Hospitalization and Oxygen Requirement Estimator: A form to be filled by the user. If in case the user doesn't have the required test result we take the mean of all the data for that particular attribute.

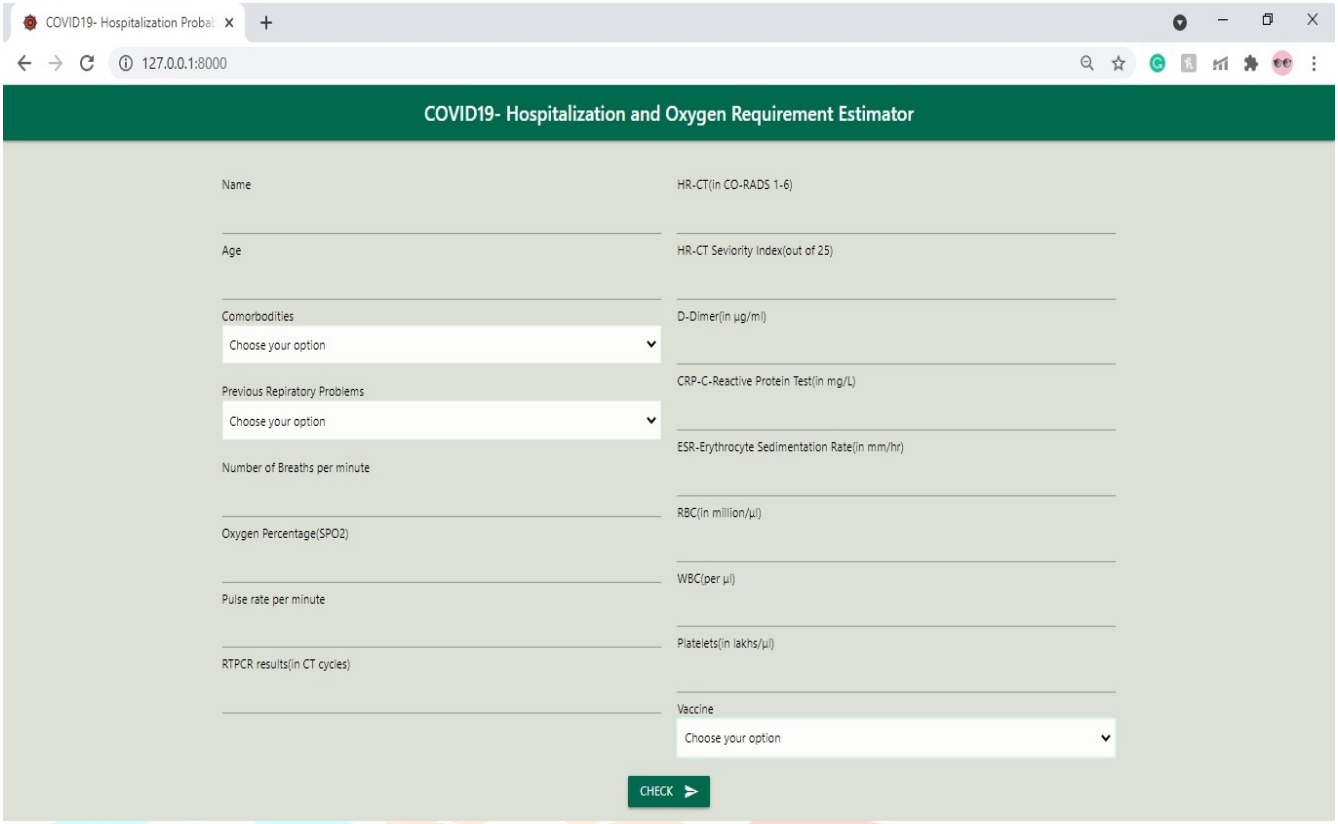

Figure 8) Webpage for COVID-19 Hospitalization and Oxygen Requirement Estimator

Result page for COVID-19 Hospitalization and Oxygen Requirement Estimator: It shows the patients details along with recommended ranges. It also gives the probability estimate of oxygen, home quarantine or hospitalization requirements. The user can print, save or share the file.

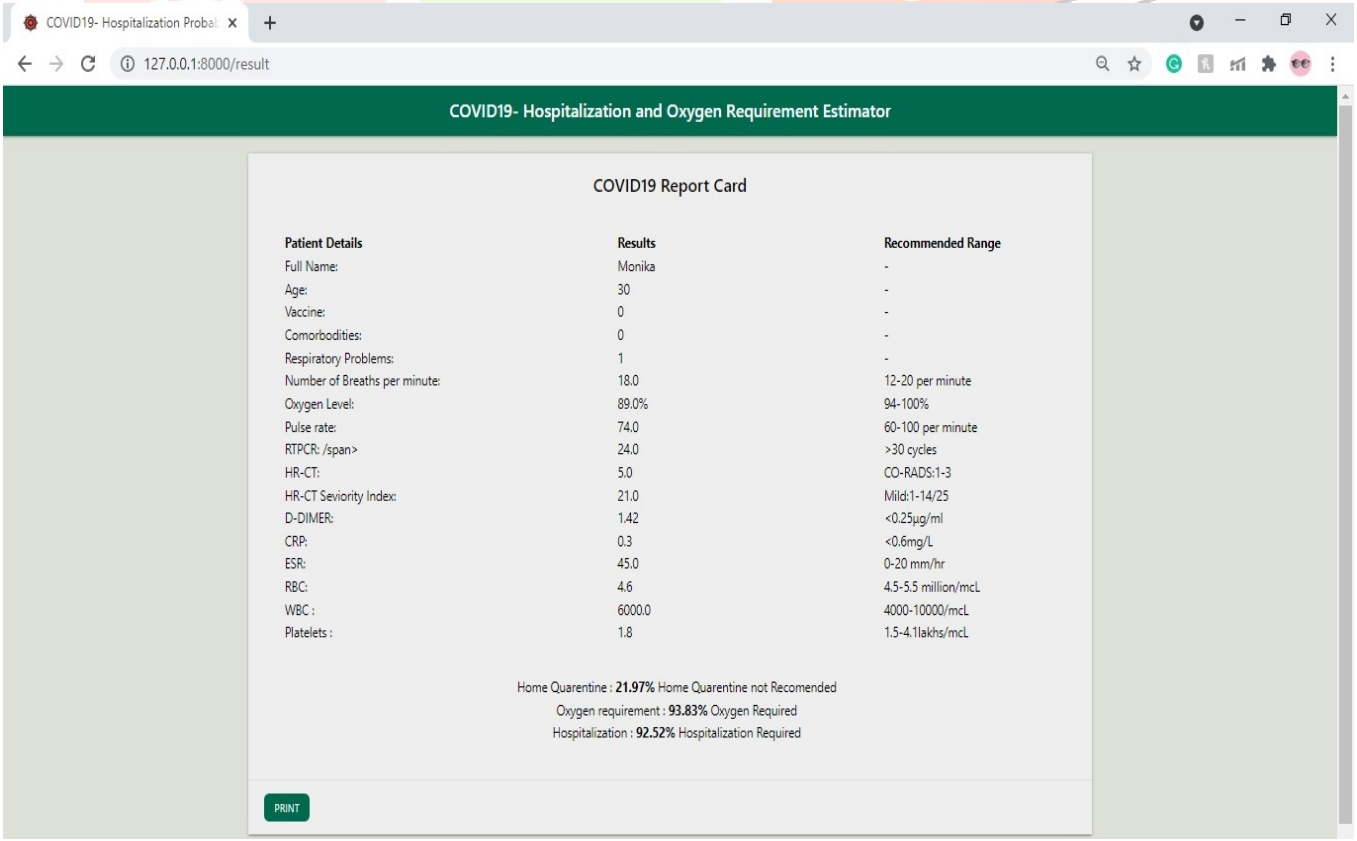

Figure 9) Result page for COVID-19 Hospitalization and Oxygen Requirement Estimator

Webpage for COVID-19 Detector: A form to be filled by the user regarding his personal details and symptoms.

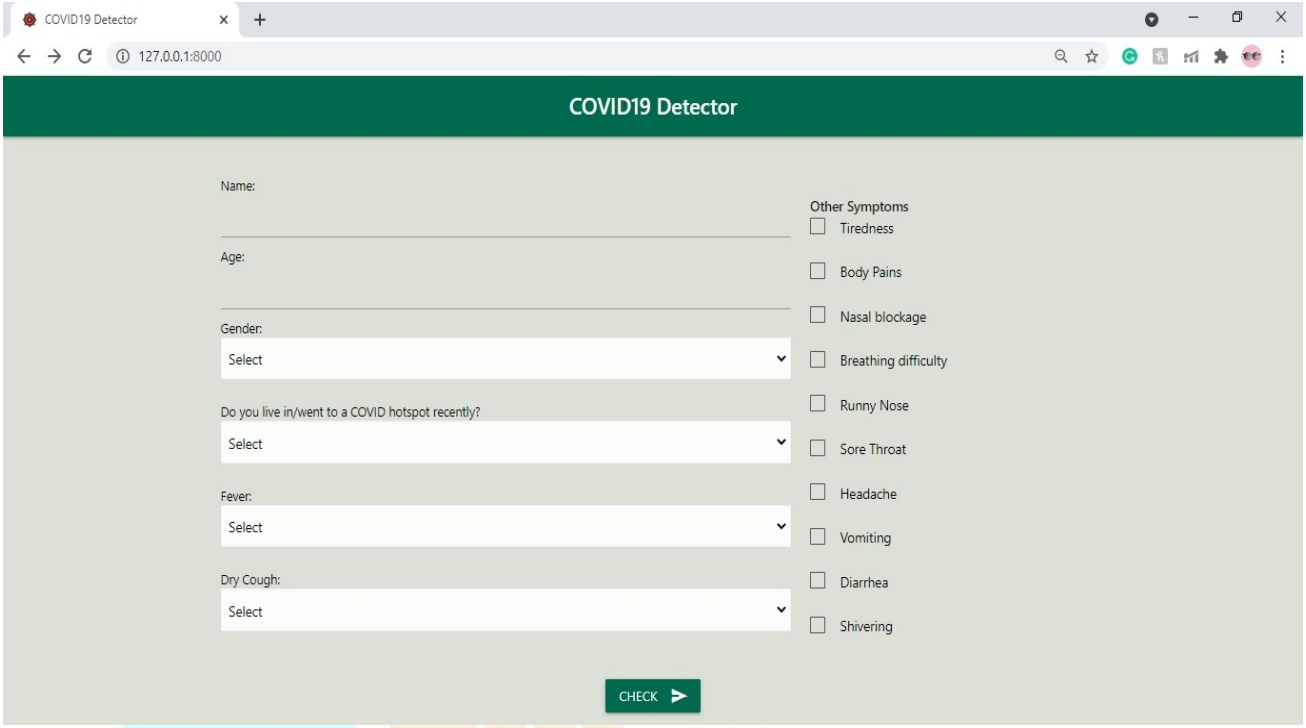

Figure 10) Webpage for COVID-19 Detector

Result page for COVID-19 Detector: It shows patients details along with probability estimate of COVID-19 infection. We can scan through the QR code and also print, save or share the file.

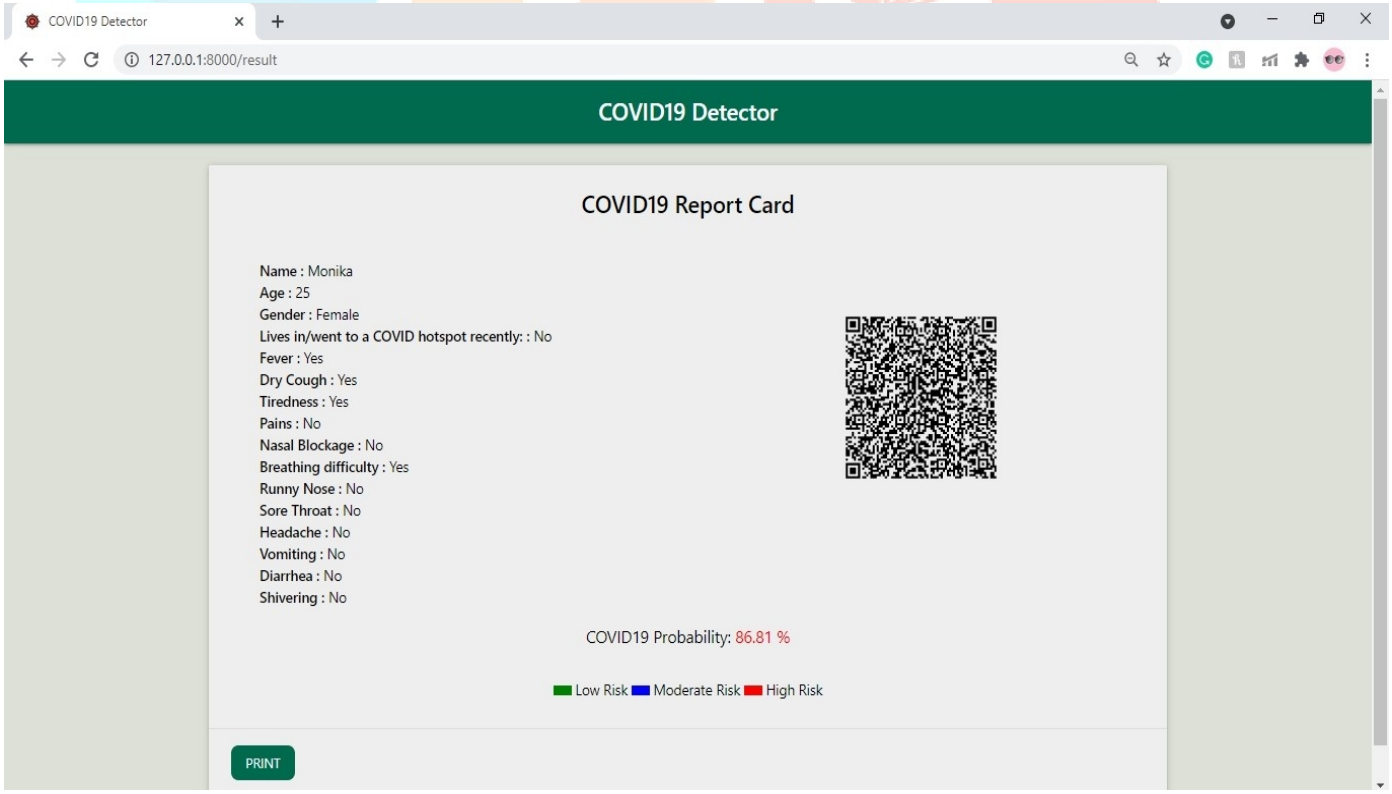

Figure 11) Result page for COVID-19 Detector

We are also able to gain insights such as, in the presence of some respiratory problems by the patient the critical oxygen requirement is a little bit lower than the normal person with no respiratory problems. In most of the cases of patients critical and requiring hospitalization, we found that factors such as D-dimer, CRP(c-reactive protein test), ESR(erythrocyte sedimentation rate) showed abnormal results. It is observed that the number of patients hospitalized after full or partial vaccination has drastically decreased when compared to non-vaccinated patients.

VII. CONCLUSION

Detecting the virus in the early days is necessary so that we can diagnose people better. This project provides us a way to estimate the probability of the coronavirus based on the symptoms of the patient by using Logistic regression with 91% accuracy. It also recommends if there is any requirement of oxygen or hospitalization based on the details of the test reports. The patient can also share their reports to doctors or health care personnel without any contact thereby decreasing the chances of COVID-19 contamination.

VIII. REFERENCES

[1] In July 2020,Thomas Struyf,Jonathan J Deeks,Jacqueline Dinnes,Yemisi Takwoingi,Clare Davenport -"Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease".https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD013665/full

[2] In Aug 2020,Gökçen Budak, Oğuz Karabay, Ertuğrul Güçlü, Hüseyin Doğuş Okan, Aslı Vatan - "Main symptoms in patients presenting in the COVID-19 period"https://journals.sagepub.com /doi/full/10.1177/0036933020949253

[3] In June 2020,Sonia Shah DO, Kaushal Majmudar DO, Amy Stein PhD, Nita Gupta MD, Spencer Suppes DO, Marina Karamanis DO, Joseph Capannari DO, Sanjay Sethi MD, Christine Patte DO proposed "Dynamic relationship between D-dimer and COVID-19 severity".https://onlinelibrary.wiley.com/doi/full/10.1111/bjh.16811

[4] In June 2020,Sonia Shah DO, Kaushal Majmudar DO, Amy Stein PhD, Nita Gupta MD, Spencer Suppes DO, Marina Karamanis DO, Joseph Capannari DO, Sanjay Sethi MD, Christine Patte DO proposed Novel Use of Home Pulse Oximetry Monitoring in COVID-19 Patients Discharged From the Emergency Department Identifies Need for Hospitalization. https://onlinelibrary.wiley.com/doi/full/10.1111/acem.14053

[5] In September 2020,Bhanu Kanth Manne , Frederik Denorme , Elizabeth A. Middleton , Irina Portier , Jesse W. Rowley , Chris Stubben , Aaron C. Petrey , Neal D. Tolley , Li Guo , Mark Cody , Andrew S. Weyrich , Christian C. Yost , Matthew T. Rondina , Robert A. Campbell has worked on COVID-19 concerns aggregate around platelets.https://ashpublications.org/blood/article/136/11/1221/463625/COVID-19- concerns-aggregate-around-platelets

[6] In the year 2020,Rea G ,Valente T , Lieto R , Bocchini G , Marchiori E 2 , Pinto A 3 , Maglio A 4 , Vatrella A -"The Many Faces Of Covid-19: Organizing Pneumonia (Op) Pattern Hrct Features".https://tmj.unisa.it/journal/vol23/iss4/2/

[7] Patient Medical Data for Novel Coronavirus COVID-19 https://datarepository.wolframcloud.com/resources/Patient-Medical-Data-for-Novel- Coronavirus-COVID-19.

[8] In March 2021,Bodunrin Osikomaiya, Olufemi Erinoso, Kikelomo Ololade Wright, Aina Olufemi Odusola, Babafemi Thomas, Oluwatosin Adeyemi, Abimbola Bowale, Olusola Adejumo, Ayodeji Falana, Ismail Abdus-salam, Olusegun Ogboye, Akin Osibogun & Akin Abayomi -"'Long COVID': persistent COVID-19 symptoms in survivors managed in Lagos State, Nigeria". https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-020-05716-x

[9] In July 2020, Nicola Magnavita , D.Giovanni Tripepi and Reparata Rosa Di Prinzio-"Symptoms in Health Care Workers during the COVID-19 Epidemic".https://www.mdpi.com/1660-4601/17/14/5218

[10] In March 2020,Hoang C. Nguyen , Minh H. Nguyen , Binh N. Do ,Cuong Q. Tran -"People with Suspected COVID-19 Symptoms Were More Likely Depressed and Had Lower Health-Related Quality of Life: The Potential Benefit of Health Literacy".https://www.mdpi.com/2077-0383/9/4/965

[11] In April 2020. Cristina Menni, Ana M Valdes, Maxim B Freidin, Sajaysurya Ganesh, Julia S El-Sayed Moustafa-"Loss of smell and taste in combination with other symptoms is a strong predictor of COVID-19 infection".https://www.medrxiv.org/content/10.1101/2020.04.05.20048421v1