Indian Health Care System is Ready to Fight Against COVID-19 A Machine Learning Tool for Forecast the Number of Beds

Shakti Nagpal
Department of Computer Science & Engineering
Panipat Institute of Engineering & Technology
Panipat, India

Vijay Anant Athavale
Department of Computer Science & Engineering
Panipat Institute of Engineering & Technology
Panipat, India

Ashwini Kumar Saini
Govind Ballabh Pant Institute of Engineering & Technology
Pauri Garhwal,

Ravindra Sharma
Swami Rama Himalayan University
Dehradun, India

2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC) | 978-1-7281-7132-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/PDGC50313.2020.9315825

***Abstract*—Global research team has announced that the health a management system at world level is in fear from CoV-19. Various statistical analysis has been done to check the preparedness to fight against CoV-19. Recent government responses of the different countries are also taken into the consideration while working for CoV-19 handling. Demographic trends are also added to add further content to potential impact of CoV-19 on healthcare services and system. This pandemic has raised a significant challenge to the economy of the different countries. Availability of beds are calculated on Per thousand people in different countries. Few of the countries analysis like Australia is having 2.6 beds per thousand people, while United Kingdom America is having 2.5 beds preparation over 1000 people. Per capita health spending in UK is marginally below the median. Hospital have been urged by government of different countries to postpone their surgeries and other treatments to provide the proper hospitality to cov-19 patients. India is at 145th place among 195 countries in healthcare access and Quality Index (HAQ)[1]. In this paper we have proposed a machine Learning model to predict the number of beds required as Cov-19 cases are increasing. Our Model Predicts the requirement for beds with 95% accuracy and acceptable p-value.**

Keywords—Machine learning, Regression, Hospitals,CoV-19, Health care Expenditure

I. INTRODUCTION

Health care system is affected all over the world due to CoV-19, or we can say it has generated an alarm to the health care services and countries are trying to devising coping strategies for it. Table 1 shows, number of beds available per thousand persons, number of doctors over thousand people, total health care expenditure done by government and the population over and above 60 years for some of the developed countries. None of the developed country was prepared and able to handle the impact of CoV-19. In this paper the statistics of the number of hospitals available in India and number of beds available for the infected patients are retrieved from Kaggle. State wise populations and number of CoV-19 patients of India is retrieved from website of the ministry of health and family welfare. Multiple Linear regression model is used for machine learning and is applied on the dataset to predict the number of required beds for the patients. An automated model is designed which can give the predictions of beds on the any entered number of patients.

Indian health care system is not robust enough to provide a quality infrastructure and other healthcare resources to millions of its citizens. CoV-19 has created an emergency to fix all bottlenecks as soon as possible. The virus has wreaked havoc in almost all the developed countries. In India it is staring at an alarming situation, just like a timebomb is fixed on the head. Due to the insufficient health care services, very low number of testing of CoV-19 has been done. Hence it is highly likely that data available in public domain may not represent the true picture of CoV-19 cases in India. For first few months Indian Council of Medical Research provided the statics on daily basis but it stopped updating. 27th March 2020, morning. After that states have been releasing figures on daily basis. These figures are also doubted by several experts as state of testing is still highly inadequate considering the population of the country. So far various State Government as well as government of India has maintained that there is no community transmission in India till date [2-4].

As per the ICMR report till now there is no community transmission of virus but the situation can change drastically overnight if it happens. Governments, panchayats and all the empowered agencies are not able to make quarantine arrangements for such a large population. Indian government has allocated Rs 15,000 crore for healthcare improvement.

TABLE I. HEALTH CARE SERVICE RECORD OF SOME COUNTRIES.(FIGURES ARE GIVEN ON PER THOUSAND PEOPLE) [SOURCE-WWW. HEALTHAFFIARS.ORG]

S.No	Health System	Australia	United Kingdom	Germany	Canada	USA
1	Health care Expenditure	7854	2,332	4,924	3,201	10,612
2	Doctors availability /thousand people	3.94	2.8	4.3	2.6	2.6
3	Beds availability	3.9	2.5	6	2.	2.8
4	Above 60 population	21.4	23.8	28.2	24.2	17.0

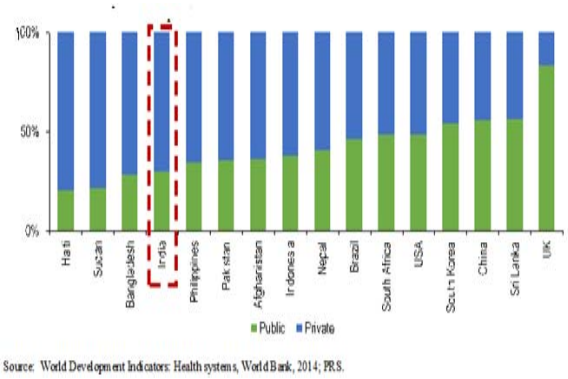
As per the national health profile-2019 data 0.55 beds per thousand population is available in India. Twelve states namely Bihar, Jharkhand, Gujarat, Uttar Pradesh, Andhra Pradesh, Chhattisgarh, Madhya Pradesh, Haryana, Maharashtra, Odisha, Assam and Manipur lie below the national figure (0.55) in terms of availability of beds over

thousand people although 70% of the Indian population lives in these states. Only some of the states like Delhi, Tamil Naidu and Kerala have the better conditions of health care system and have 1.05 beds availability per thousand persons. It is obvious that number of beds available per thousand populations in India is very low. Health care service records of some countries are shown in table 1.

Sudden outbreak of coronavirus pandemics has complicated the problem, the priorities of health care systems has been changed around the world. Limited availability of resources and time the inefficiencies in the health care systems are magnified.

The demand of healthcare services has suddenly increased. CoV-19 patients are struggling hard to get access to the health care facilities and beds in the hospitals, In India we have 7,13, 896 number of government hospitals beds and 5-10% ICU beds facility is available. Only 50% of these ICU beds have ventilators, and other emergency health care services available. As per the report published in times of India 17,850 to 25,556 ventilators in the country.

Total expenditure on health care services in India is approximately 267 $ per capita and health percentage of GDP is 4.7 while the life expectancy of Indians are 69.17, the public sector healthcare expenditure is 53bUSI.



Source: World Development Indicators: Health systems, World Bank, 2014; PRS.

Fig. 1. Heath care expenditure of various countries [5]

In figure 1 the report is published by the world health organization where expenditure done by individual on health care system is depicted. In India approximately 70% of the health care expenses are done by the personal pockets which leads to push 7% of the population to below the poverty line. The public health expenditure in India is constant during 2008 to 2015 which is approximately 1.3% of GDP and marginally increased to 1.4% in 2016-2017 but still less than 6% of the total world average. While the National health policy proposed a rise to 2.5% of GDP by 2025.

TABLE II. HAQ INDEX OF DIFFERENT COUNTRIES [6]

HAQ Rankings For BRICS			
Country	HAQ Index Score, 2016	HAQ Index Rank, 2016	Improvement In Score, 1990-2016
India	41.2	145	16.5
Brazil	63.8	96	17.3
Russia	75.1	58	11.9
China	77.9	48	35.3
South Africa	49.7	127	9.6

Source: Global Burden of Disease study published in the *Lancet*

In table 2 HAQ (health care access and quality Index) index of different countries is mentioned which was reviewed in 2016 and published the same in the report of times of India in April 2020. Where India is at 145th positions. HAQ index is provide on the basis of global burden of diseases which includes approximately 32 causes of deaths which are preventable and curable through effective medical care.

Two of the states in India like GOA and Kerala have highest HAQ index and reached to 60 points whereas Assam and UP have lowest score i.e. below than 40.

II. OUR CONTRIBUTION

In this paper we have developed a machine learning model which will helps us to forecast the number of beds required for the CoV-19 patients. As on July 5, 2020, India has 6,83,000 coronavirus cases which is increasing at very faster rate. The Indian CoV-19 positivity rate is 6.73 percent. There is major responsibility goes with the government and health ministry to take proactive measures to handle the situation. The proposed model will dynamically provide the number of hospital and beds for taking care of the patient. Linear regression model of machine learning is used to implement the forecast problem. Linear regression is the best machine learning model for the continuous data flow and provides the accuracy of prediction after proper training of the model. It is a supervised machine learning model where dataset of different states of hospitals, beds, population of states as well as the number of CoV-19 patients per state is given as input. The data set is divided into two phases training set and testing data into 70 -30 ratio [7-11].

The following Objectives have been covered and implemented-

- Processing of population of Indian states, number of hospitals, number of beds available in hospitals and number of CoV-19 infected patients.
- Applied machine Learning algorithm for processing the data and making the predictions of required number of beds for the CoV-19 patients
- Designed a Generalized graphical model for better predictions of number of beds required on the basis of number of patients.

III. MULTIPLE LINEAR REGRESSION

Multiple Linear regression is the most common form of linear regression analysis, which is used to signify the relationship between single continuous dependent variable and more than one independent variables. In this paper number of beds re continuous variable which is dependent upon the number of hospitals and number of infected cases of CoV-19 [14-18].

Steps

1. Get the two data sets with n examples and let us denoted it as x and y for each.

X -is taken as number of cases and hospitals

Y- the number of beds available

2. Find out the variable m by using the equation

$$m = \frac{(\sum_i x_i y_i) - n\bar{x}\bar{y}}{(\sum_i x_i^2) - n\bar{x}^2}.$$

3. Then find out the variable b by using the equation,

$$b = \bar{y} - m\bar{x}.$$

4. At last substitute the value of b and min the linear regression equation

$$f(x) = mx + b$$

Where the applied models generate the value of m which is intercept

Results and Discussion

import pandas as pd

import matplotlib.pyplot as plt

df=pd.read_csv ('c:\\users\\shakt\\Desktop\\current.csv')

A. *Dataset*

The dataset for the availability of number of beds and hospitals available in India was accessed from Kaggle named "Hospitals and beds in India (state wise)" as depicted in figure 2. The dataset of the state wise populations and number of CoV-19 infected cases is also compiled from the "Ministry of health and family welfare" portal.

df.info

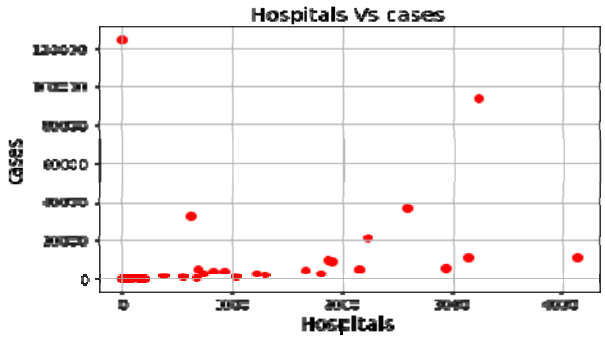


Fig. 2. Hospitals Vs Cases

<bound method DataFrame.info of

0	Andaman & Nicobar Islands	380581	34	34	1246
1	Andhra Pradesh	49577103	1666	5269	60799
2	Arunachal Pradesh	1383727	199	57	2320
3	Assam	31205576	1220	3092	19115
4	Bihar	104099452	2146	5710	17796
5	Chandigarh	1055450	47	327	3756
6	Chhattisgarh	25545198	1023	1262	14354
7	Dadra & Nagar Haveli	585764	13	26	568
8	Daman & Diu	16787941	8	2	298
9	Delhi	1458545	615	32810	20572
10	Goa	60439692	40	387	2666
11	Gujarat	25351462	2236	21521	41129
12	Haryana	6864602	683	5579	13841
13	Himachal Pradesh	12267032	671	451	8706
14	Jammu & Kashmir	32988134	818	4507	11342
15	Jharkhand	61095297	558	1489	7404
16	Karnataka	33406061	2943	6041	56333
17	Kerala	274000	1297	2161	39511
18	Lakshadweep	64473	10	115	250
19	Madhya Pradesh	72626809	1867	10049	38140
20	Maharashtra	112374333	3239	94041	68998
21	Manipur	2570390	114	311	2562
22	Meghalaya	2966889	180	44	4585
23	Mizoram	1097206	87	93	2312
24	Nagaland	1978502	166	128	1944
25	Odisha	41974218	1799	3250	16497
26	Puducherry	1247953	53	127	4462
27	Punjab	27743338	742	2805	13527

Fig. 3. Sample of Data used for predictions

B. *Data Analysis*

Figure 3- Sample of Data used for predictions

Figure 4- Resultant Parameters of the Applied

Figure 5- Correlation between selected attribute

Number of hospitals, availability of beds per hospital, population of the individual state and number of CoV-19 infected cases are the important parameters considered for further analysis and predictions.

intercept: m = [1038.59763481]

Coefficients are the Constant b1 and b2 used to

Coefficients: - [[0.0724152 , 17.77107088]]

we can predict any value of y when X as input.

New _ cases = 100000

New _ hospitals = 50

OLS Regression Results

Dep. Variable:	Beds	R-squared:	0.825
Model:	OLS	Adj. R-squared:	0.815
Method:	Least Squares	F-statistic:	82.54
Date:	Thu, 18 Jun 2020	Prob (F-statistic):	5.62e-14
Time:	16:49:39	Log-Likelihood:	-400.98
No. Observations:	38	AIC:	808.0
Df Residuals:	35	BIC:	812.9
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1038.5976	2131.337	0.487	0.629	-3288.247	5365.442
Cases	0.0724	0.065	1.120	0.270	-0.059	0.204
hospitals	17.7711	1.465	12.131	0.000	14.797	20.745

Omnibus:	11.709	Durbin-Watson:	2.100
Prob(Omnibus):	0.003	Jarque-Bera (JB):	15.541
Skew:	0.825	Prob(JB):	0.000422
Kurtosis:	5.663	Cond. No.	3.70e+04

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.7e+04. This might indicate that there are strong multicollinearity or other numerical problems.

Fig. 4. OLS Regression results

print ('Predicted Beds: \n', regr.predict ([[New_cases , New_hospitals]]))

Predicted Beds:

[[9168.67090395]]

C. *Correlations*

All the crucial information about the necessary features and the degree of influence over the target value can be retrieved from the correlation between the features. The correlation between the number of beds and total number of cases for specific period is depicted in fig. as well as the number of hospital available corresponding to the number of cases in Fig 1 and Fig.2 which reveals the strong relationship and a positive correlation between the cases and beds as well as number of hospitals vs cases.

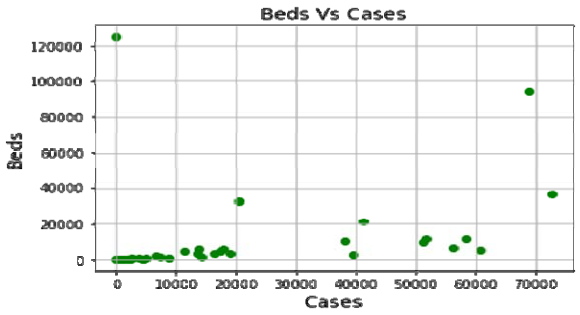


Fig. 5. Correlation between selected attribute

D. *Data Pre-processing*

The data set consist of various columns with datatype being string, date and Numeric. In machine learning all input data should be in numerical format, labelled encoding is applied on the categorical data of the columns which assign a unique number to each categorical data in the column.

Data wrangling is applied on the dataset to fill the missing values with NAN, in some of the states where hospitals are mentioned but the number of beds is missing are filled with the average number of beds available in states. The column having a data type date could not be used directly to handle the date feature engineering is applied on the specific column. The complete dataset is divided into two phases- training data and Test data. 70% of the dataset is used to train the machine while 30% of the dataset is used to test the data.

E. *Evaluation*

The objective of the evaluation metrics is to accurately predict the outcome of number of beds required as per the number of infected cases increased. The accuracy of the results is very import in the figure 6. where correlation is positive, standard error is 0.0724, predicted value is acceptable 0.629 which is greater than 0.5 proves that proposed model is acceptable. The dataset consists of positive and negative datapoints, and the accuracy of the model could be calculated by total number of data points divided by total number of correct predictions divided by (positive plus negative). With the help of accuracy, the performance of the designed system can be analysed accuracy of the implemented model could be calculated with the following equation.

accuracy= Total data points/ total number correct predictions

$$0.0 < \text{Accuracy} < 1.0$$

While precision of the system is the ratio of positive samples divided by total number of datapoints. Precision helps in identifying the correct number of accurate/correct predictions.

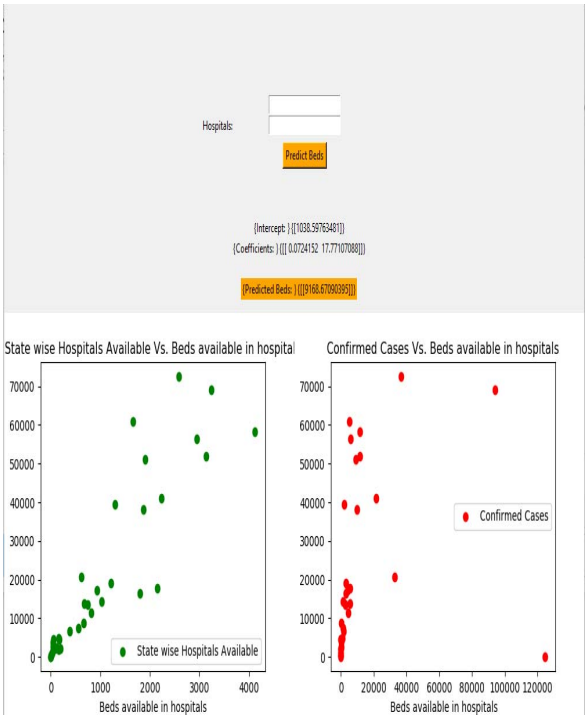


Fig. 6. Designed Model Demonstration

IV. CONCLUSION

A machine learning model is applied on the provided dataset, where multiple linear regression model is applied on the various attributes. State-wise population vs hospitals and beds with respect to number of infected cases of CoV-19 are considered for predicting the futuristic requirements. The prediction is acceptable with 95% accuracy. An automated interface of the proposed model is designed which can dynamically provide the predictions on the desired dataset (number of Cov-19 cases which are continuously changing). The proposed model is designed with limited dataset available on different portals but the information is inconsistent. The major issues regarding the usage of the predicted outcome is the lack of authentic data available. Government has stopped publishing the exact cases, testing and status of the CoV-19 on public portals. Machine learning models always provides the optimized results after the successful training with accurate dataset.

REFERENCES

[1] https://www.statista.com/statistics/1117473/india-health-risk-assessment-by-gender.

[2] https://www.statista.com/topics/5191/state-of-health-in-india.

[3] https://www.businesstoday.in/current/economy-politics/coronavirus-lockdown-covid-19-pandemic-public-healthcare-system-doctors-nurses-patients/story/400039.html

[4] https://currentaffairs.gktoday.in/tags/haq-index

[5] https://openknowledge.worldbank.org/bitstream/handle/10986/18237/9781464801631.pdf

[6] George J. Schieber, Jean-Pierre Poullier, and Leslie M. Greenwal, health care systems in twenty-four countries, *Health Affairs*, Vol.10, no.3,pp-21-38 Available online at https://www.healthaffairs.org/doi/full/10.1377

[7] https://www.brookings.edu/blog/up-front/2020/03/24/is-indias-health-infrastructure-equipped-to-handle-an-epidemic.

[8] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. (2020) 395:497–506. doi: 10.1016/S0140-6736(20)30183-5

[9] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N Engl J Med. (2020) 382:1199–207. doi: 10.1056/NEJMoa2001316

[10] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. (2020) 395:507–13. doi: 10.1016/S0140-6736(20)30211-7

[11] Clinical Management of Severe Acute Respiratory Infection When Novel Coronavirus (2019-nCoV). Infection Is Suspected: Interim Guidance. (2020). Available online at: https://apps.who.int/iris/handle/10665/330893 (accessed April 31, 2020).

[12] Zumla A, Hui DS, Perlman S. Middle East respiratory syndrome. Lancet. (2015) 386:995–1007. doi: 10.1016/S0140-6736(15)60454-8.

[13] Pham QV, Nguyen DC, Hwang WJ, Pathirana PN. Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: a survey on the state-of-the-arts. Preprints. (2020) 2020:2020040383.doi: 10.20944/preprints202004.0383.v1

[14] WHO Situation Report-94 Coronavirus disease 2019 (COVID-19). (2020). Available online at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200423-sitrep-94-covid-19.pdf?sfvrsn=b8304bf0_4 (accessed March 10, 2020).

[15] Kathiresan S, Sait ARW, Gupta D, Lakshmanaprabu SK, Khanna A, Pandey HM. Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. Pattern Recogn Lett. (2020) 133:210–6. doi: 10.1016/j.patrec.2020.02.026

[16] Wang L, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. arXiv. (2020) 2003.09871. Available online at: https://arxiv.org/abs/2003.09871 (accessed May 5, 2020).

[17] Pal R, Sekh AA, Kar S, Prasad DK. Neural network-based country wise risk prediction of COVID-19. arXiv. (2020) 2004.00959. Available online at: https://arxiv.org/abs/2004.00959 (accessed May 7, 2020).

[18] Liu D, Clemente L, Poirier C, Ding X, Chinazzi M, Davis JT, et al. A machine learning methodology for real-time forecasting of the 2019–2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. arXiv. (2020) 2004.04019. Available online at: https://arxiv.org/abs/2004.04019 (accessed May 6, 2020).