Predicting the Probability of Covid-19 Recovered in South Asian Countries Based on Healthy Diet Pattern Using a Machine Learning Approach

Md. Showrov Hossen [1], Dip Karmoker [2]
Faculty of Computer Science and Engineering
Patuakhali Science and Technology University
Patuakhali, Bangladesh
sourovhossen96@gmail.com[1] , karmokerdip75@gmail.com[2]

Abstract— **Recently a global terror has taken place around all over the world named of COVID-19 disease. The main cause of this disease is SARS-CoV-2 virus. A huge number of population of the world is losing lives daily for this terrible virus. But in the global survey for this disease we found that the people also getting recovered from this frightening disease. The most important thing that works behind the recovery against this virus is the immunity power of human body. Immunity power is not same for all human body. Immunity power of human depends on the food habit of them. In this research we will try to determine the probability of COVID-19 recovered in South Asian Countries based on healthy diet pattern using data mining and various machine learning algorithms. We have used Random Forest, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) are the several machine learning algorithms to predict the recovery rate of Covid-19 affecting patients.**

Keywords— COVID-19, SARS-CoV-2, Coronavirus, healthy diet, immunity

I. INTRODUCTION

COVID-19 disease is the name of a frightening panic at the current situation of the whole world. COVID-19 is the recent form of Coronavirus. This virus is known as SARS-CoV-2 virus (Severe Acute Respiratory Syndrome Coronavirus 2). COVID-19 first spread around the region of Hubei in China in late 2019 [1] [2]. Enormous lives claimed for this. After China this virus has been take place to almost every country. As it is a contagious disease everyday a huge amount of people are getting affected with this virus and expires. But due to good immunity power some affected patients get rid of this deadly disorder. Immune system of a human body generally depends on the healthy food habit. This immunity system is playing an important role in the case of COVID-19. This research is aimed to determine the probability of recovery from the COVID-19 in South Asian countries through analyzing the diet pattern of this region.

II. MOTIVATION

Analyzing current situation of COVID-19 disease and some concerned research paper, we become motivated to work on this topic and that is prediction of the recovery probability of COVID-19 in South Asian countries based on healthy eating style. Some significant issues noticed by us not analyzed yet:

- Comparison of the diet pattern of South Asian countries and top 10 COVID-19 affected countries.

- Prediction of the affected patient recovery rating from COVID-19 based on healthy diet.

III. LITERATURE REVIEW

SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) is a recent form of Coronavirus which first started spreading in the state of Hubei in China. As a result of the outbreak of this terrible virus a large amount of human lives were claimed [3] [4] [5]. According to the declaration of the WHO in January 2020, the outbreak of novel coronavirus is considered as a Public Health Emergency of International Concern (PHEIC) [6] [7]. After that again in February 2020, WHO declared an official name of this contagious disease, COVID-19 (Coronavirus Disease 2019). Then this disease COVID-19 declared as a Pandemic in March 2020 [7] [8].

Generally Coronavirus is a family of viruses. This virus causes various infections and Respiratory tract diseases. This can be enormous in various cases like SARS, MERS, and COVID-19 [1]. There are various types of coronavirus that affect animals. As some rare cases, this virus outstretched from brute species into the human population [9]. From a current study it's come to know that if the coronavirus outbreak starts once, it will take less than four weeks to enwrap the total healthcare system. If the hospital capacity gets enwrapped, the death rate will jumps [10].

COVID-19 generally causes severe respiratory tract infections which is the leading reason of mortality and morbidity around the whole world [11]. To reduce the expansion and dominance of these respiratory viruses some public health practices are founded like wearing masks, maintaining social distance, washing hands, covering coughs

with masks. [12]. These rules are currently applying around the globe as a protection till an effective drug is not developed against COVID-19. But all of these are not enough. Our immune system plays an important role as the central player that fight against this coronavirus. It helps us to be protected from this invader virus. This is also helpful for its therapy [13]. A good immune system is extremely important to combat against COVID-19. Maintaining a good immunity, sufficient vitamins and other nutrients are badly needed for the body. But many people are lack of this proper vitamins and nutrients [14]. Maintaining a potential immune system a healthy diet plays an important role and this assists to preventing diseases. Sustaining a nutritious diet and healthy lifestyle is highly significant during the COVID-19 pandemic.

All the things mentioned above are about the background knowledge of COVID-19 disease and the role of human immunity system in this case. Healthy diet and lifestyle is the key of good immunity in human body. This knowledge has assist us to gather all the ideas that we exert in our research work. With the help of this gathered knowledge we have made our research work more organized and a better analysis of the features that are present in our dataset.
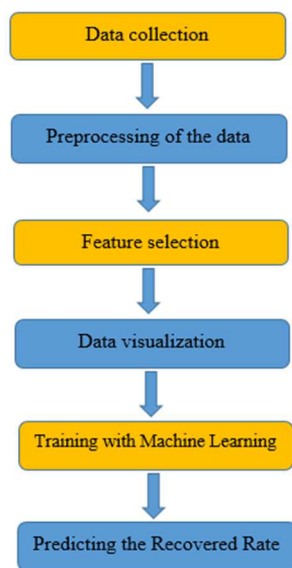
IV. System Architecture

Fig. 1. Methodology Diagram

In this research work our methodology is proceeded after scraping data set of Covid-19 Healthy Diet Dataset. Then the term preprocessing is to remove trash values from the dataset. After that the selection of valid data will be performed. Now analyzing all the dataset features will be visualized from various aspect. Afterward the selected data will be trained to machine learning and using different machine learning algorithm the recovery rate will be predicted.

V. Data Preprocessing

Usually data arrives with garbage values which should be removed so that it cannot be able to affect the performance of trained models that predict the outcome.

A. Datasets

We have considered here Covid-19 Healthy Diet Dataset to perform the research work. The source of this dataset is kaggle.com

Country	Alcoholic Beverages	Animal Products	Cereals - Excluding Beer	Meat	Vegetal Products	Others	Recovered	Death	Unit
Afghanistan	0	4.7774	37.1186	1.2006	45.2476	11.6558	82	18	%
Albania	0.912	16.093	16.2107	3.8688	33.907	29.0085	96	4	%
Algeria	0.0896	6.0326	25.0112	1.2543	43.9749	23.6374	87	13	%
Angola	1.9388	4.6927	18.3521	2.9302	45.3184	26.7678	93	7	%
Antigua and Ba	2.3041	15.3672	13.7215	7.0356	34.6225	26.9491	91	9	%
Argentina	1.4354	14.9869	16.7927	9.4459	34.99	22.3491	94	6	%
Armenia	0.2274	12.833	19.2658	4.2235	37.167	26.2833	97	3	%
Australia	1.9783	15.6146	11.643	7.4902	34.3854	28.8885	95	5	%
Austria	2.8161	15.6106	12.3748	5.0095	34.3894	29.7996	98	2	%
Azerbaijan	2.2555	8.1682	28.7417	2.9322	41.8237	16.0787	98	2	%
Bahamas	2.9145	14.793	10.7274	8.1803	35.2437	28.1411	78	22	%
Bangladesh	0	2.3695	37.5265	0.3468	47.6401	12.1171	93	7	%
Barbados	1.1224	12.0877	15.3514	5.664	37.9209	27.8536	85	15	%
Belarus	3.2979	12.2644	14.2705	5.7143	37.7204	26.7325	98	2	%

Fig. 2. Sample Dataset

B. Feature Selection

Executing our research work we have selected some features from the collected dataset and they are Country, Alcoholic Beverages, Animal Products, Cereals - Excluding Beer, Meat, Vegetal Products, Others, Recovered, Death.

VI. Exploratory data analysis

In this section we have visualized the comparison of calories consumption from different food products between South Asian countries and top 10 Covid-19 affecting countries. Top 10 Covid-19 affecting countries are selected from the report of the website Worldometer coronavirus update April 2020.

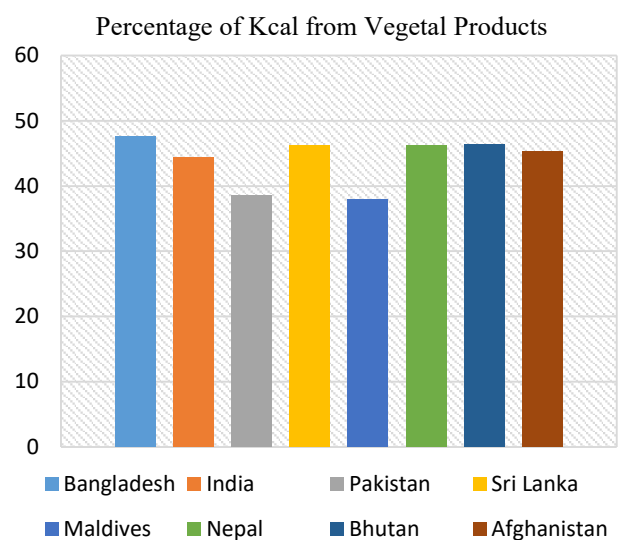### A. Calories Consumption from Vegetal Products in South Asian Countries

Fig. 3. Percentage of Kcal from Vegetal Products for South Asian Countries

Calories Consumption from Vegetal Products in Top 10 Covid-19 Affecting Countries
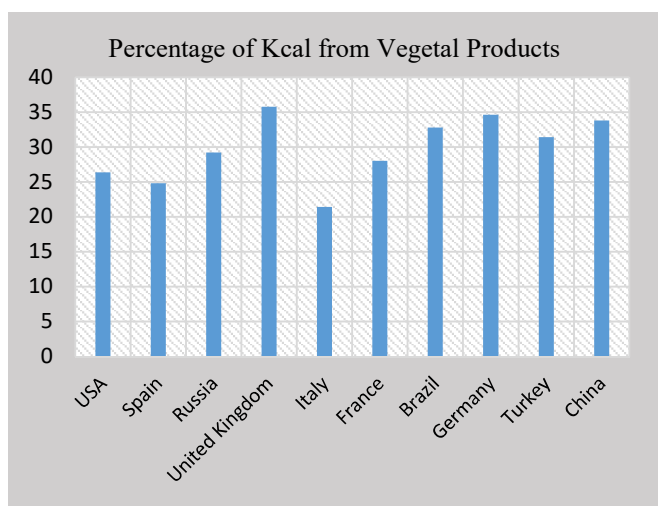

Fig. 4. Percentage of Kcal from Vegetal Products for Top 10 Covid-19 Affecting Countries

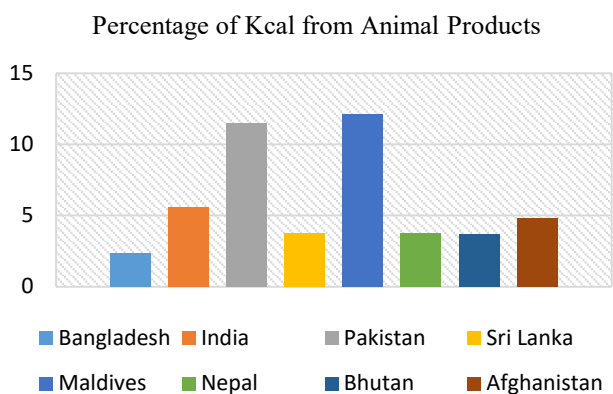B. *Calories Consumption from Animal Products in South Asian Countries*

Fig. 5. Percentage of Kcal from Animal Products for South Asian Countries

Calories Consumption from Animal Products in Top 10 Covid-19 Affecting Countries
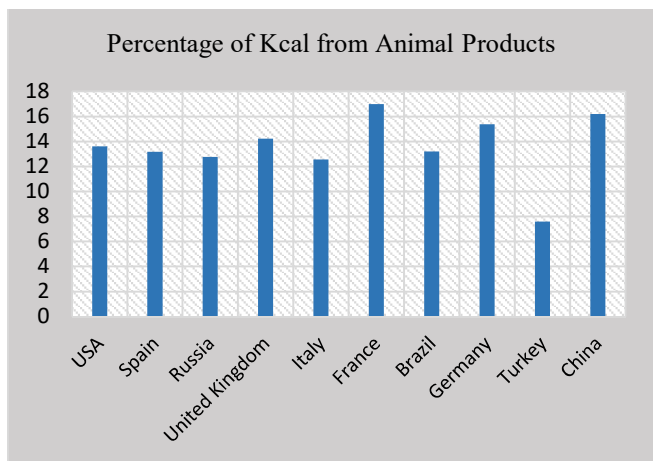

Fig. 6. Percentage of Kcal from Animal Products for Top 10 Covid-19 Affecting Countries

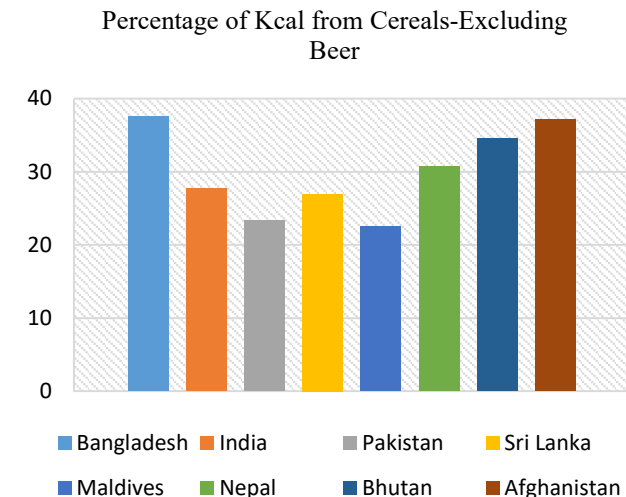C. *Calories Consumption from Cereals-Excluding Beer in South Asian Countries*

Fig.7. Percentage of Kcal from Cereals-Excluding Beer for South Asian Countries

Calories Consumption from Cereals-Excluding Beer in Top 10 Covid-19 Affecting Countries
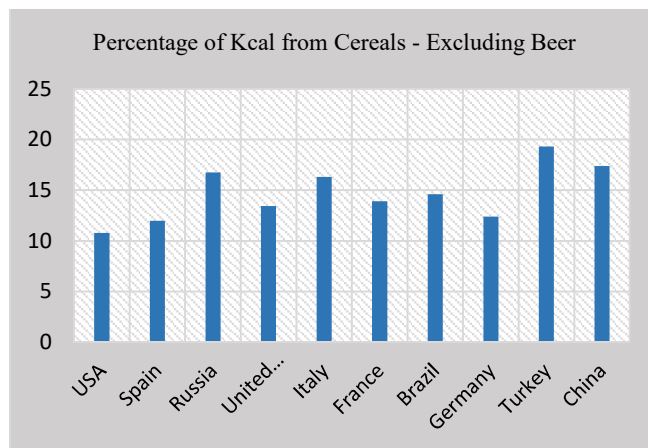

Fig. 8. Percentage of Kcal from *Cereals-Excluding Beer* for Top 10 Covid-19 Affecting Countries

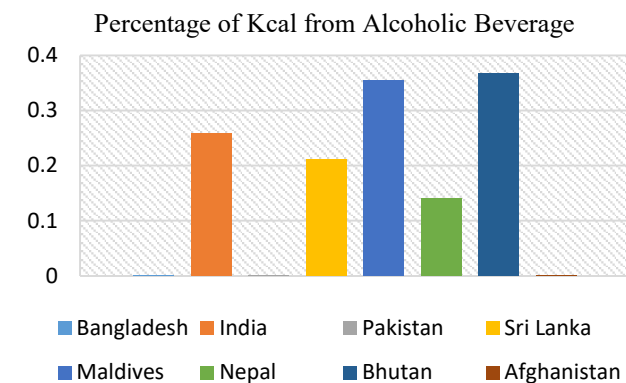D. *Calories Consumption from Alcoholic Beverage in South Asian Countries*

Fig.9. Percentage of Kcal from Alcoholic Beverage for South Asian Countries

Calories Consumption from Alcoholic Beverage in Top 10 Covid-19 Affecting Countries
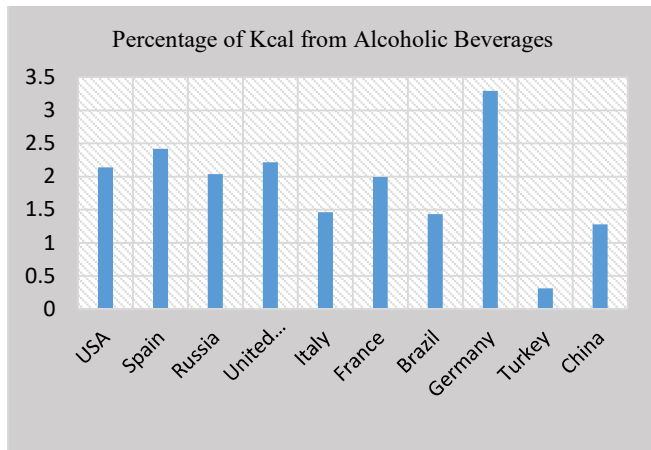

Fig. 10. Percentage of Kcal from *Alcoholic Beverage* for Top 10 Covid-19 Affecting Countries

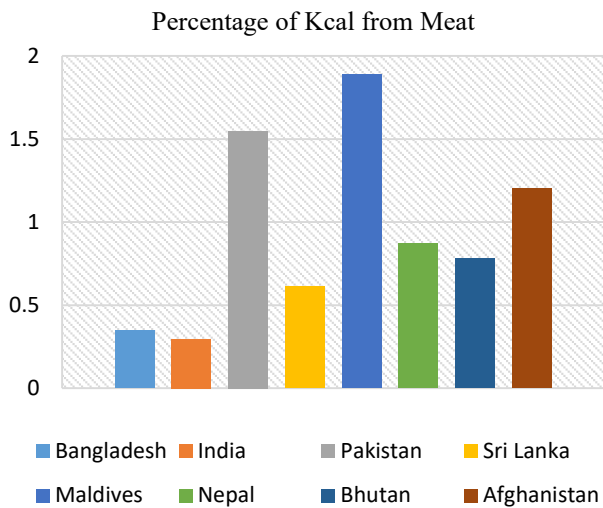*E. Calories Consumption from Meat in South Asian Countries*

Fig.11. Percentage of Kcal from Meat for South Asian Countries

Calories Consumption from Meat *in Top 10 Covid-19 Affecting Countries*
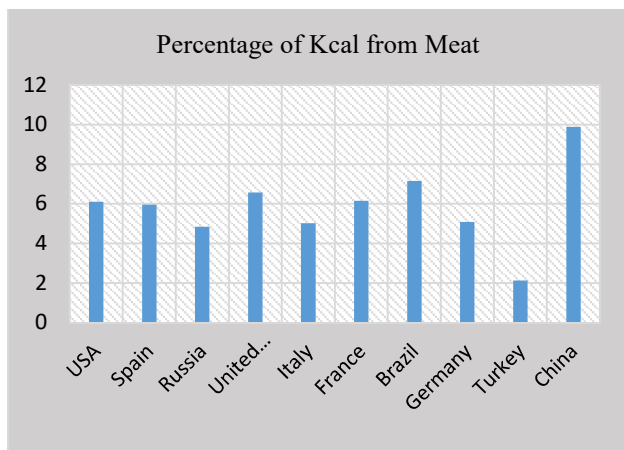

Fig. 12. Percentage of Kcal from Meat for Top 10 Covid-19 Affecting Countries

VII. RESULT ANALYSIS

Analyzing the result here we have considered 10 case (Patients) recovered percentage prediction from Covid-19 by using 3 different types of machine learning algorithm. Three Machine Learning Algorithms – Random Forest, KNN and SVM are used for the prediction of recovery rate.
Some used short forms are in the result table discipted here.

Input Attributes:
   Percentage of Kcal Consumption From-A.B, A.P, C.E.B, M, V.P, Ot.
A.B-   Alcoholic Beverages
A.P-   Animal Products
C.E.B- Cereals Excluding Beer
M-   Meat
V.P-   Vegetal Products
Ot-   Others
Output Attributes:
   Predicting the Percentage of Covid-19 Recovered Rate-Rec. P
Rec. P- Recovered Prediction

Result by Used Algorithms

   Random Forest: Random forest algorithm is generally a tree based algorithm. Operating this algorithm it needs to build various decision trees and prepare the combination of their output to improve deduction ability of the model. Here is the procedure of Random Forest algorithm is

- Random selection of "k" features from entire "m" features (Alcoholic Beverages, Animal Products, Cereals Excluding Beer, Meat, Vegetal Products, Others) Where k << m
- Calculation of the node "d" from the "k" features using the best split point.
- Division of the node into daughter nodes by using the best split.
- Repetition of 1st to 3rd steps until "l" number of nodes has been reached.
- Creating "n" number of trees to build the forest repeat steps 1st to 4th for "n" number times.

TABLE I: Predicted result of recovery rate by Random Forest algorithm

Test	A.B %	A.P %	C.E.B %	M %	V.P %	Ot %	Rec.P %
Case 1	0	2	37	2	47	12	**94**
Case 2	1	4	32	3	42	18	**92**
Case 3	1	3	35	3	44	14	**93**
Case 4	0	3	39	2	48	08	**96**
Case 5	1	5	33	5	41	15	**90**
Case 6	4	15	24	8	35	14	**81**
Case 7	6	17	29	5	33	10	**79**
Case 8	5	12	27	4	39	13	**82**
Case 9	3	19	25	6	34	13	**80**

Case 10	4	20	26	5	31	14	**77**

KNN: KNN (k- nearest neighbor) algorithm is used to classify an unknown sample based on it's neighbors known classification. A sample of unknown classification is predicted by the consideration of the of its nearest neighbor samples classification. Here is the procedure of KNN algorithm is

- Classify(X: training data, Y: class labels of X, x: unknown sample)
- for i =1 to m do
- Computation of distance d(Xi, x)
- end for
- Computation of set I which contain indices for the k smallest distances d(Xi, x)
- Return majority label for{Yi where i ∈ I}

TABLE II: Predicted result of recovery rate by K-Nearest Neighbor algorithm

Test	A.B %	A.P %	C.E.B %	M %	V.P %	Ot %	Rec.P %
Case 1	0	2	37	2	47	12	**91**
Case 2	1	4	32	3	42	18	**93**
Case 3	1	3	35	3	44	14	**90**
Case 4	0	3	39	2	48	08	**92**
Case 5	1	5	33	5	41	15	**94**
Case 6	4	15	24	8	35	14	**81**
Case 7	6	17	29	5	33	10	**77**
Case 8	5	12	27	4	39	13	**80**
Case 9	3	19	25	6	34	13	**78**
Case 10	4	20	26	5	31	14	**81**

SVM: In the case of linearly separable data in two dimensions a machine learning algorithm tries to find a boundary to reduce the misclassification error. Decision boundary is chosen in SVM (support vector machine) algorithm. Then the distance from the nearest data points of all the classes is maximized by it. It determines the most favorable decision boundary.

In the case of nonlinearly separable data, the simple SVM algorithm doesn't work. Then Kernel SVM, is used. Here is the procedure of SVM algorithm is

- Initialization of k solutions
- Call SVM algorithm for the evaluation of k solutions
- T= Sort $(S_I......S_k)$
  While classification accuracy != 100 or number of iteration != 10 do
- for i= 1 to m do
- select S according to its weight
- sample selected S
- store newly generated solutions
- call SVM algorithm the evaluation of newly generated solutions
- end for
- T= Best (Sort $S_{I,....}S_k$ + m),k)
- end while

TABLE III: Predicted result of recovery rate by Support Vector Machine algorithm

Test	A.B %	A.P %	C.E.B %	M %	V.P %	Ot %	Rec. P %
Case 1	0	2	37	2	47	12	**92**
Case 2	1	4	32	3	42	18	**91**
Case 3	1	3	35	3	44	14	**94**
Case 4	0	3	39	2	48	08	**93**
Case 5	1	5	33	5	41	15	**89**
Case 6	4	15	24	8	35	14	**78**
Case 7	6	17	29	5	33	10	**80**
Case 8	5	12	27	4	39	13	**81**
Case 9	3	19	25	6	34	13	**82**
Case 10	4	20	26	5	31	14	**79**

VIII. CONCLUSION

COVID-19 disease is a name of terror around the whole world at the current time. There is no vaccine or antidote available till now to everyone. But we can be able to battle with Covid-19 by adapting a healthy diet. A healthy eating style could help to combat the Corona Virus. We have made an analysis on the energy intake from different categories of food of different countries. Predicting the percentage of recovery rate of COVID-19 various machine learning algorithms are used like Random Forest, SVM, KNN where the input attributes are percentage of Kcal consumption from Alcoholic Beverages, Animal Products, Cereals Excluding Beer, Meat, Vegetal Products and Others food habit of a patient. We have shown the bar chart of calories consumption from various types of food in South Asian Countries. We also visualized through bar charts the calories consumption from various types of food at top Covid-19 affected countries. According to the analysis we have observed that top Covid-19 affected countries on average consumes more animal products than South Asian countries while consuming more plant based products(e.g. cereals, vegetables). By different bar charts we compared here the diet pattern of South Asian countries and top 10 Covid-19 affected countries. Through this research work we have come to a decision that when a patient consumes more plant based products (e.g. cereals, vegetables) and less animal products then recovered percentage is more otherwise less. Our proposed methodology will help the people of South Asian countries to predict the recovery probability in early stage based on healthy dieting style.

IX. REFERENCES

[1] M. S. Mohammad Pourhomayoun, "Predicting Mortality Risk in Patients with COVID-19 Using Artificial Intelligence to Help Medical Decision-Making," medRxiv, Los Angeles, 2020.

[2] T. T. Nguyen, "Artificial Intelligence in the Battle against Coronavirus (COVID-19): A Survey and Future Research Directions," preprint, 2020.

[3] H. F. D. K. H. E. L. Todd Ellerin, "Harvard Health Publishing Coronavirus Resource," 2020. [Online]. Available: https://www.health.harvard.edu/diseases-and-conditions/coronavirus-resource-center.

[4] Li, Qun & Guan, Xuhua & Wu, Peng & Wang, Xiaoye & Zhou, Lei & Tong, Yeqing & Ren, Ruiqi & Leung, Kathy & Lau, Eric & Wong, Jessica Y & Xing, Xuesen & Xiang, Nijuan & Wu, Yang & Li, Chao & Chen, Qi & Li, Dan & Liu, Tian & Zhao, Jing & Li, Man & Feng, Zijian. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine. 382. 10.1056/NEJMoa2001316.

[5] Xu, B., Gutierrez, B., Mekaru, S. *et al.* Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci Data* **7,** 106 (2020). https://doi.org/10.1038/s41597-020-0448-0.

[6] Isaac I Bogoch, Alexander Watts, Andrea Thomas-Bachli, Carmen Huber, Moritz U G Kraemer, Kamran Khan, Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel, *Journal of Travel Medicine*, Volume 27, Issue 2, March 2020, taaa008. https://doi.org/10.1093/jtm/taaa008.

[7] S. o. t. s. m. o. t. I. H. Regulations, "World Health Organization," WHO, 30 January 2020. [Online]. Available: https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov).

[8] W. D.-G. o. r. a. t. m. b. o. COVID-19, "World Health Organization (WHO) (Press release)," 11 March 2020. [Online]. Available: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

[9] M. M. Lisa Maragakis, "Johns Hopkins Medicine," 15 March 2020. [Online]. Available: https://coronavirus.jhu.edu/.

[10] B. P. E. A. Trent McConghy, "When does Hospital Capacity Get Overwhelmed in USA? Germany?," Medium, 2020.

[11] Calder, P.C.; Carr, A.C.; Gombart, A.F.; Eggersdorfer, M. Optimal Nutritional Status for a Well-Functioning Immune System Is an Important Factor to Protect against Viral Infections. *Nutrients* 2020, *12*, 1181.

[12] U. C. f. D. C. T. 3. A. t. F. Flu, "Centers for Disease Control and Prevention," National Center for Immunization and Respiratory Diseases (NCIRD), 17 October 2019. [Online]. Available: https://www.cdc.gov/flu/prevent/preventing.htm.

[13] A. Carthaus, "DW," 7 April 2020. [Online]. Available: https://p.dw.com/p/3aaRK.

[14] J. Vergin, "DW," 07 May 2020. [Online]. Available: https://p.dw.com/p/3btZW.