International Journal of Humanities and Social Science Invention (IJHSSI)
ISSN (Online): 2319 – 7722, ISSN (Print): 2319 – 7714
www.ijhssi.org ||Volume 11 Issue 10 October. 2022 || PP. 90-102

Measuring the Risk of COVID-19 Pandemic in West Bengal Districts

Souvik Dey[1]
[1]*Department of Economics, The University of Burdwan*
Burdwan, West Bengal, India, Pin - 713104.

ABSTRACT:*To realise the risk of COVID-19 in different countries several models have been evolved in the last two years. Construction of a social vulnerability index and exploring the effect of natural disasters on social vulnerability is an area of interest for researchers. In this paper, we have tried to merge the social vulnerability by considering several significant factors of the novel COVID-19 pandemic. We have selected a total of 17 indicators related to COVID-19 status and socio-economic aspects. After normalizing the selected variables, we have used the most dynamic method of index construction the Principal Component Analysis (PCA) to get the final Index of COVID Vulnerability (CVI). The seventeen dimensions have been reduced in two independent principal components and we haveruled out the other components using the Kaiser criteria of their respective Eigenvalues. The study reflects the ranking of West Bengal districts in terms of their respective CVI scores from highly vulnerable areas to less vulnerable regions. After getting the CVI scores we have successfully tested the existence of spatial dependency by Moran's I correlation coefficient. Some districts have been discovered to be grouped with high levels of COVID vulnerability, whereas some other districts are found to be adjacent with lower levels of vulnerability. In the conclusion, we have included that the developed areas have a higher risk of COVID-19 as compared to less developed districts.*

KEYWORDS:*Principal Component Analysis; COVID Vulnerability Index (CVI); Spatial autocorrelation; Moran's I correlation coefficient.*

Date of Submission: 12-10-2022 Date of Acceptance: 28-10-2022

I. INTRODUCTION

Even though it has been almost two years since the timeof 2019, we are still fighting against the world's one of the most life-threateningdiseases, the Corona Virus pandemic. The first case of novel coronavirus was found in the Huanan Seafood Wholesale market of Wuhan City, China. The Chinese government decided on December 31, 2019, to notify the World Health Organization (WHO) about an unknown illness that causes respiratory difficulties in humans such as high fever, dry cough, tiredness, and other symptoms. Within a week after the 31st of December, more than 44 people in China were diagnosed with severe pneumonia caused by an unknown virus. After that,on 30th January 2020,the World Health Organization (WHO) avowed the COVID-19 prevalence as Public Health Emergency of International Concern (PHEIC[1]). Within the first two weeks of February 2020, this unknown virus was discovered outside of China in 25 additional countries. More than 70,000 people outside of China have been infected with the coronavirus, while 1,772 people have died in China. On March 11, 2020, the World Health Organization proclaimed the coronavirus to be a global pandemic.The number of confirmed cases doubles every week in the early stages of the virus's spread in other nations. The WHO announced in the middle of October 2020 that there were over 35 million confirmed cases throughout the world. The symptoms of the coronavirus are very common including the loss of smell & taste, headache, runny nose and fever. Though the symptoms are varied from person to person the most common and frequent three indicationsof this virus are identified as respiratory issues, shortness of breath and fever. The worst feature of the novel coronavirus is that it is contagious which makes the severity of the virus more brutal.

Every year, a large number of individuals come from other countries to China (and vice versa) for tourism, employment, education, or other reasons and India has no exception for it. The risk of infectious diseases like COVID in a huge, overcrowded nation like India with less developed healthcare resources is quite significant, and the threat should be treated more severely than in other developed countries. On January 30, 2020, the first confirmed case of coronavirus in India was discovered in a 20-year-old female in Kerala, India. Due to the outbreak of COVID in China, she was reported to be travelling from Wuhan, China, to India. After

---

[1]*World Health Organization. Novel coronavirus (2019-nCoV) situation report – 11. Geneva: WHO; 2020. Jan 31.*

that, the infection spreads throughout all Indian states, with more than half of all confirmed cases concentrated in only six states including Kolkata, Delhi, Mumbai, Pune, Ahmedabad, and Chennai. The first countrywide lockdown to control the spreading of the virus was announced by the Indian Government starting from 24th March 2020 to 1st June 2020. As predicted, the intensity of the virus's spread in India exceeded the United States record on August 30, 2020, just two months after the initial lockdown period ended, and according to news sources, the record was more than 78,000 COVID positive cases in a single day. On September 16, 2020, India set a new record with more than 98,000 confirmed cases in a single day. On the Asian continent, India is the country with the highest number of confirmed cases in 2020, and it becomes the world's second-most severe country after the United States in September 2020. After the first wave of COVID-19 in India,the Government of India has published that the cases in India have crossed the 10 million number on December 2020 but it was also stated that the pace of spreading become slower than earlier after the lockdown periods. The healthcare services in India have been severely affected by a second wave of COVID-19 between March to April 2021.So we have decided to explore the condition of coronavirus severity in the Indian state of West Bengal till the latest month of February 2021.

The first positive case of COVID-19 in West Bengal was found in a male student who travelled from England to Kolkata on 15th March 2020. Within March, a total of 20 persons including 11 years old children to 57 years old persons have reported to be COVID positive in different districts of West Bengal. At the beginning of the outbreak, the majority of positive patients in West Bengal districts were reported to have travelled from other countries to West Bengal. On 7th April 2020, the West Bengal state Government disclosed the 7 hotspot areas including Howrah, Kolkata, Haldia, Belgharia, Tehatta, Egra and Kalimpong where the people have been instructed to stay home, stay safe.Due to the risk of spreading COVID-19 in the districts, the West Bengal State Government has announced on 14th March 2020 to close all the schools, colleges and universities till the 31st March 2020 and next it was decided to close the educational institutions to 15th April of the same year. As a result, 7.9-crore individuals who get subsidised rations are directed to work from home. The government would also give assurance to provide free rations to the underprivileged till September 2020. There is a directive from the government of West Bengal to all districts to provide temporary housing and food for migrants and the poor. For the impact of COVID-19, the West Bengal state government was setting up an Rs. 200-crore fund on 23rd March 2020. 8 districts in West Bengal are among the 170 COVID-19 hotspot districts that have been designated by the central government. There are 4 red zones in West Bengal including Kolkata and 348 containment zones[2]. The status ofCOVID-19 cases and other statistics as on February 2021 from the earlier stage of spreading the threat are shown in the following table.

Table 1: Status of COVID-19 in West Bengal Districts till February, 2021

Districts	Total Positive Cases of COVID-19	Total Discharged Persons from COVID-19	Total Deaths from COVID-19	Total Active Cases
Kolkata	127931	123790	3080	1163
North 24-Parganas	121989	118287	2486	1216
South 24-Parganas	37002	36024	712	266
Howrah	35564	34335	1043	186
Hooghly	29461	28795	496	170
Burdwan	28857	28186	267	404
Paschim Medinipur	23216	22728	331	157
Nadia	22524	22046	317	161
Jalpaiguri	22309	21949	247	143
Purba Medinipur	20613	20128	279	206
Darjeeling	20557	20136	229	192
Malda	12655	12453	114	88
Murshidabad	12264	11990	149	125
Bankura	11873	11616	92	165
Coochbehar	11820	11686	72	62
Birbhum	9979	9759	89	131

[2] *"District Wise Containment Zones by Government of West Bengal". West Bengal State Portal. Retrieved on 16th May, 2020.*

Dakshin Dinajpur	8161	8044	74	43
Puruliya	7186	7021	48	117
Uttar Dinajpur	6623	6457	74	92
West Bengal	**570584**	**555430**	**10199**	**5087**

Source: 2021 nCoV Bulletin as on 2nd February, 2021, Health & Family Welfare Department, The Government of West Bengal. (Sorted in descending order by the number of positive cases)

In this paper, we are focusing on the following objectives,

1. To construct the COVID Vulnerability Index (CVI) using the Principal Component Analysis to measure the risk or condition of COVID-19 in West Bengal districts till the year 2021.
2. To explore the severity of the districts by focusing on any spatial dependency among the districts or in other words we want to study the existence of spatial autocorrelation among the districts regarding their COVID Vulnerability Index (CVI).

II. REVIEW OF LITERATURES

There are several existing kinds of literature for understanding the vulnerable situation of districts level and states level in India after facing the pandemic situation. ***Sarkar and Chouhan*(2021)** explained in their paper that the pandemic has posed a significant threat to the world, as well as to Indian society and economy which has turned into a public health crisis, with its severity rising all the time. The foremost objective of the paper is to construct a socio-environmental vulnerability index of the potential risk of community spread of COVID-19 with the help of some socio-economic and environmental variables. They have used principal component analysis to construct the ultimate vulnerability index of the districts of India. After the empirical analysis, the findings demonstrate that vulnerability varies spatially depending on environmental and socioeconomic factors, and districts of north and central India established more vulnerable than south India.

Through the paper of ***Buvinic, Noe,and Swanson*** **(2020)**, we can understand the situation of women's and girls' vulnerabilities to the COVID-19 pandemic. They highlighted the condition of the women and girls of 75 lower and middle-income group countries around the world. The chief purpose of the paper is to identify the most vulnerable LICs and MICs where women and girls are at the highest risk of suffering primary and secondary health effects of pandemic and also to make the ranking of the countries based on each index of girls' and women's wellbeing. To understand women's wellbeing, three individual indexes like women's health, economic opportunities, and human capital are constructed and also finally assembled a composite women's vulnerability index. This paper also prescribed the policy that the gender vulnerability data dashboard is one of the important instruments for monitoring the situation of girls and women in vulnerable countries.

***Tavares and Betti*(2020)** described in their paper that how poor individuals in Brazil in terms of indicators are directly connected to their ability to avoid and recover from COVID-19 infection, as well as how much and in what ways they are disadvantaged. To assess multidimensional poverty in the context of the coronavirus pandemic, they combined the Alkire-Foster (AF) technique with a fuzzy-set methodology. This paper also tried to establish two pandemic-specific indices to account for the vulnerability associated with the ability to prevent and recover from illness infection. According to rank correlation studies, the suggested indices can track trends in rising infection and increased mortality in susceptible areas. Their empirical findings suggested that pandemic responses must target the most vulnerable individuals and emphasizes the need for national coordination.

***Brito, Kuffer, Koeva, et al*(2020)** highlighted in their paper that in comparison to the formal city, where residents have the resources to fulfil WHO recommendations, impoverished communities' living circumstances, and urban morphologies render them more vulnerable to the COVID-19 epidemic. They explored that municipal spatial databases are unprepared to assist spatial responses to health emergencies, analysed especially in low-income areas. So they suggested that Earth Observation (EO) data aid in rapid decision-making and perhaps save many lives. This study offers an indication of the possibilities of EO-based global and local datasets, as well as local data collection techniques in support of COVID-19 responses by mentioning two slum areas in Salvador, Brazil as a case study.They utilized the example of two regions where the group has developed long-standing research on infectious illnesses and collaboration with community leaders to assess the unique geographical data needs for impoverished communities in the COVID-19 location**.**

Barraza, Barrientos, Díaz, et al **(2020)** scrutinized that the entire cost of human life has yet to be determined which has been lost under the Covid-19 pandemic situation. They also explored that along with the high cost of life and a serious health crisis, the globe is experiencing an economic downturn that will have a significant influence on the well-being of broad segments of the people in the next years. In this paper, the Multidimensional Poverty Index is used to identify pre-existing poverty circumstances that make certain families more vulnerable to the pandemic than others (MPI). This paper also explored that MPI identifies six factors that contribute to these risks such as access to drinking water, health services, overcrowding, sanitation, underemployment, and social security. One of the aforementioned deprivations is projected to affect 85.5 percent of families.

Zhenlong, Xiaoming, Dwayne, et al **(2020)** explained in their paper about the human movement which is one of the important forces that drives the geographical spread of infectious illnesses. During the COVID-19 pandemic, restricting and tracking human mobility has been shown to help limit the virus's transmission. This research seeks to create a unique data-driven public health strategy that monitors and evaluates human movement at various spatial scales by combining big data from Twitter with other human mobility data sources and artificial intelligence (from global to regional to local). They first generate a database with optimal spatiotemporal indexing to tackle the multisource data sets. Then, utilizing geo-tagged big data from Twitter and other human mobility data sources, they created innovative data models, prediction models, and computer algorithms to efficiently extract and analyse human movement patterns.

Poom, Jarv, et al **(2020)** discussed the individual movement that has been disrupted to unprecedented levels by the COVID-19 epidemic. They explained that the crisis is spatial inspecting the geographical aspect which is very important under the pandemic situation. The rush of mobile Big Data enables researchers to investigate the crisis' geographical consequences in spatiotemporal depth at national and global dimensions. They suggested two strategic pathways for the future use of mobile Big Data for societal impact evaluation, covering both raw and aggregated mobile Big Data products. They also concluded that both pathways need careful consideration of privacy concerns, standardized and transparent procedures, and a focus on data emblematic dependability, and continuity.

III. METHODOLOGY & DATA

This study is based on several indicators which can explain the condition of COVID-19 and the Socioeconomic Status of 19 West Bengal districts. To capture the status of the COVID-19 situation in West Bengal districts we have managed to collect the secondary data from the West Bengal covid-19 health bulletin, CoWin dashboard of Ministry of Health & Family Welfare, Integrated Covid Management System of West Bengal (West Bengal-ICMS) till the time February 2021. And the other socio-economic data have been collected from the Census of India, West Bengal state statistical handbook, etc. After collecting the data we have used the SPSS version 25 statistical software to run the Principal Component Analysis and constructing the COVID Vulnerability Index (CVI).

Our focus is to construct the COVID Vulnerability Index (CVI) of West Bengal districts using the available data till February 2021. There have been a variety of ways to create the index throughout the past few years. The social vulnerability index was created by Cutter. Et al. (2003), which is a comparative statistic that offers an indication of an area's social vulnerabilities to various threats. Through a method known as principal components analysis, the index is generated by synthesizing socioeconomic factors. The factors used to build the index were chosen after considerable research on socio-economic and geographical parameters that can affect the area economically as well as geographically. So here in our study, we have used the approach of Principal Component Analysis (PCA) to reduce the dimensions and compute the final index of COVID vulnerability. To create a composite index score using the PCA, we must first normalise the indicators using the UNDP's *"Goal Post"* approach, which has been widely accepted for decades. Before creating the COVID Vulnerability Index, we must first choose certain indicators that can significantly explain the state of coronavirus vulnerability considering the COVID-related and socio-economic aspects. We have selected a total of 17 indicators that are divided into two parts to construct the final composite index of COVID vulnerability. The selected indicators are mentioned in the following table.

Table 2: Selected Indicators of COVID Vulnerability Index (CVI)

Dimensions	Indicator's Definition
COVID-related Indicators	Total Positive Cases of COVID-19
	Total Discharged Persons after COVID-19
	Total Deaths from COVID-19
	Total Active Cases
	Percentage of supply of PPE Kits
	Supply of Total N95 Masks
	Numbers of COVID-19 Special Government Hospitals
	Availability of Government Hospital Beds for COVID Patients
	Availability of Safe Home Beds for COVID Patients
	Numbers of vaccination sites

	Numbers of the total vaccinated population
Socio-economic Indicators	Population Density
	Percentage of population below 7 years age*
	Percentage of the population more than 65 years old*
	Literacy Rate
	Per capita Income
	Number of total SC and ST population*

Source: Author's selection
**Estimated values for the year 2021 based on Census 2011*

After collecting the data according to our selected indicators mentioned above for West Bengal districts till the year 2021, we have applied the UNDP's *"Goal Post"* method of normalization to transform them into a unique unitary value. The method of standardization uses the formula of,

$$I_{i,j} = \frac{In_{i,j} - Min(In_{i,k})}{Max(In_{i,k}) - Min(In_{i,k})}$$

Where, $I_{i,j}$ is the i^{th} indicator for the j^{th} district, $In_{i,j}$ is the actual value of the i^{th} indicator for the j^{th} district, $Min(In_{i,k})$ is the minimum value of that i^{th} indicator among all the k districts, and $Max(In_{i,k})$ is the maximum value of the i^{th} indicators among all the k districts. The normalization process will transform the individual indicators into the score between 0 to 1. A score closer to 1 implies a better situation while a value closer to 0 indicates a poorer condition in the case of index scores such as Human Development Index, Human Poverty Index, etc. Contrary to what is implied by these mentioned index calculations, the situation will be reversed if there is an index of vulnerability. That is, the Higher the value of CVI will be considered as the worse condition as the status of vulnerability is higher in those regions, on the other hand, the lower the value will be considered as a better status of vulnerability.

Indexes may suffer from strong inter-correlated dimensions or correlations within the indicators, which can arise the complexity in constructing a composite index. By using the PCA approach instead of directly using indicators, we can minimise the weighting biases and reduce inter-correlation[3] between the indicators to arrive at final index scores. The Principal Component Analysis will lead to developing a reduced number of variables than the total number of selected variables (17 in our case) which are called the Principal Components (PCs). These PCs are now uncorrelated with each other. Initially, when we do the PCA using any statistical software we will get the same numbers of PCs as the number of indicators we have. But instead of taking all the PCs, we need to follow the *Kaiser Criteria* to select the PCs, which considers only those components for which the respective Eigenvalues are greater than one. Another thing that we need to keep in mind is that these Eigenvalues are not close to our original dataset as the components are not in rotated form. So after computing the principal components we have to rotate the components the orthogonal varimax rotation to represent the actual status of the indicators. The respective rotated components are nothing but the linear function of the selected indicators which can be written as,

$$PC_1 = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$
$$PC_2 = b_1X_1 + b_2X_2 + \cdots + b_nX_n$$
$$\vdots$$
$$PC_n = n_1X_1 + n_2X_2 + \cdots + n_nX_n$$

Here $PC_1, PC_2, \ldots, PC_n$ are the rotated principal components, $a_1, b_1, \ldots, n_n$ are the coefficients of each indicator that are to be estimated, $X_1, X_2, \ldots, X_n$ are the selected indicators. So here we must get the n numbers of components as we have the n numbers of indicators. After checking the *Kaiser Criteria* we will select the PCs for which the Eigenvalues are greater than one. After getting the reduced numbers of uncorrelated components we need to put weightage to each component and these weights are nothing but ……… (1) the percentage of variation described by the respective rotated principal components. We can now estimate the final COVID Vulnerability Index (CVI) by simply adding up these weighted components. This can be explained using the following equation.

Let suppose we have '*m*' numbers of rotated components from 'n' numbers of indicators after applying the *Kaiser criteria*. Hence the final CVI can be generated using the formula of,

$$CVI_j = \sum_{i=1}^{m} (PC_{i,j} * \gamma_i) \qquad \text{……… (2)}$$

[3]*Mishra, S.K., 2007. A Comparative Study of Various Inclusive Indices and the index Constructed by the Principal Components Analysis. MPRA Paper No.3377.*

Where CVI_j is the COVID Vulnerability Index for the j^{th} district in West Bengal, $PC_{i,j}$is the i^{th} number of rotated principal components among 'm' numbers of total components and γ_i is the proportion of variation explained by the i^{th} component i.e. respective weights of the components. Higher the score of CVI represents the region is highly vulnerable to live for COVID situation while lower the value of CVI implies the district is comparatively secure to COVID-19 pandemic.

Following our second objective, we wish to investigate if there is any geographical link between CVI among the districts in West Bengal. Moran's I correlation coefficient can be used to assess the spatial autocorrelation of the data. However, before we can estimate the Moran's I coefficient, we must first construct the spatial weight matrix. Two techniques may be used to determine the Moran's I correlation coefficient, namely the global version of Moran's I as well as the local version of Moran's I.

The Moran's I statistics measures the overall spatial autocorrelation among the entities in our whole study area. It helps us to understand how one area is surrounded by the similar values of its neighbouring areas for any socio-economic parameter. If the entities are enticed to each other, which means the regions are not spatially independent. And it disrupts the basic notion of independence of observations. The presence of spatial dependency may lead to an unsuitable statistical result, so it is important to test whether our data is spatially independent or not. And Moran's I is the most accepted autocorrelation test to check spatial dependency. The Moran's I[4] coefficient is calculated using the formula mentioned in equation 3,

$$I=\frac{n\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}(X_i-\bar{X})(X_j-\bar{X})}{S^2\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}}\ (i\neq j) \qquad \text{.........} (3)$$

Where n is the number of spatial areas (19 districts in our data), X_i and $X_j$are the CVI scores of i^{th} and j^{th} districts respectively. S^2 is the sample variance calculated using the formula of, $S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ where, $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$. The term W_{ij} matrix of spatial weights consists of the elements of *i* rows and *j* columns. This weight matrix can be defined using three different ways such as whether the regions are geographically contiguous to each other, bygeographical distance between the areas, and economical distance between areas. In other words, the weight matrix is formed to define which area is the neighbour of which one and this neighbouring region can be selected using the above three mentioned senses. Here in our study, we have used the geographical contiguous weights system i.e. when two districts *i* and *j* are geographically adjacent to each other it will be considered as a neighbour of each other but it should be noted that a region cannot be the neighbour of itself. On the other hand, when two districts *i* and *j* are not geographically adjacent, it will not be considered as a neighbouring region. When the region defines as a neighbour of another region the value of W_{ij} takes 1, otherwise, it takes 0.

$$W_{ij}=\begin{cases}1, \text{ when } i^{th} \text{ district is contiguous to the } j^{th} \text{ district}\\ 0, \text{ when } i^{th} \text{ district is not contiguous to the } j^{th} \text{ district}\end{cases}$$

Where *i, j* = 1, 2… 19. Hence the weight matrix can be defined as the following form,

$$W_{ij}=\begin{pmatrix}0 & w_{12} & \dots & w_{1n}\\ w_{21} & 0 & \dots & w_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ w_{n1} & w_{n2} & \dots & 0\end{pmatrix}$$

Here 'w_{11}' takes the value 1 if district 1 is the neighbour of district 1 but as we mentioned earlier that any region cannot be a neighbour of itself, so the diagonal of the matrix is consist of zero values. Similarly 'w_{12}' will take the value 1 if district 1 is the neighbour of district 2 otherwise it takes 0. In that way, we have used the queen's contiguity method to construct the 19*x*19 (row by column) spatial weight matrix *(binary form)* for the West Bengal districts.

To study the clusters or correlation of CVI among the districts, the LISA or Local Indicators of Spatial Association investigation method may be used. It decides the degrees of correlation of CVI among spatial areas. The test statistic that used the local Moran's I coefficient is formulated as,

$$I_i=\frac{n(X_i-\bar{X})\sum_{j\neq i}W_{ij}(X_j-\bar{X})}{\sum_i X_i-\bar{X})^2}$$
$$I_i=\frac{(X_i-\bar{X})\sum_{j\neq i}W_{ij}(X_j-\bar{X})}{S^2} \qquad \text{.........} (4)$$
$$=Z_i\sum_{i\neq j}W_{ij}Z_j$$

[4]*Moran, PAP: Notes on continuous stochastic phenomena. Biometrika. 1950, 37: 17-23*

Where $Z_i = \frac{X_i - \bar{X}}{S}$ and $Z_j = \frac{X_j - \bar{X}}{S}$ are the normalized observation values, $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. W_{ij} is the spatial weight matrix as mentioned above in the global Moran's I method. The final index score of local Moran's I will be lies between -1 to +1, however, any correlation coefficient estimation takes the value from 0 to +1 i.e. no correlation to perfect correlation but in the case of Moran's I coefficient, it has minor different interpretation due to introduction of a complicated spatial control. The final index score can be differentiated and define in the following way,

$$I_i = \begin{cases} -1, \text{"High-Low" clustering or "Low-High" clustering} \\ 0, \text{ No local correlation} \\ +1, \text{"High-High" clustering or "Low-Low" clustering} \end{cases}$$

Here the “High-Low” or “Low-High” cluster means the perfect clustering of dissimilar values or we can interpret it as the perfect dispersion among the regions across the CVI. The zero index coefficient represents there is no spatial dependency among the regions. “High-High” or “Low-Low” cluster shows the perfect clustering of similar values or we can think that there is a gathering of districts with high CVI scores or low CVI scores.

IV. Empirical Findings

Before discussing the results of our analysis let us first confirm the sample data adequacy to check which tool would be appropriate for our study between Factor Analysis and the PCA. We have checked the Kaiser-Meyer-Ohlin (KMO) test statistics to determine the sample data adequacy and Bartlett’s Test of Sphericity to test the existence of inter-correlation among the selected variables.

Table 3: Results of Kaiser-Mayer-Ohlin (KMO) Test & Bartlett’s Test

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.644
Bartlett's Test of Sphericity	Approx. Chi-Square	609.65
	Df	136
	Sig.	0.000

Source: Based on the author’s calculation

The value of KMO test statistics is always ranged between 0 to 1, where the value closer to 1 (or greater than 0.6) implies that the selected samples are significantly adequate and convened to the components. Here in this study, the KMO test statistic is found as 0.644 which is greater than the suggested benchmark of sample adequacy. So it indicates that our selected indicators are suitable to perform the PCA. Bartlett’s Test of Sphericity is needed to check whether the correlation matrix is an identity matrix or not. Identity matrix refers to the matrix where the diagonal elements of the matrix are 1 and other non-diagonal elements are 0 i.e. there is no correlation among the variables. The null hypothesis of Bartlett’s test is the correlation matrix is an identity matrix and we need to reject the null hypothesis (at $p<0.05$). Here in our result, the approx. Chi-Square value is found as 609.65 (at 136 degrees of freedom and Sig. level of 0.000<0.05), which is highly significant to reject the null hypothesis. Hence from the above two test results, we can say that our sample is adequate and also there is a correlation among the variables therefore the Principal Component Analysis will be the suitable procedure to construct the index scores.

Table 4: Extracted Communalities from PCA for the selected indicators of CVI

Variables	Initial	Extraction
Total Positive Cases of COVID-19 (X_1)	1	0.925
Total Discharged Persons after COVID-19 (X_2)	1	0.925
Total Deaths from COVID-19 (X_3)	1	0.914
Total Active Cases (X_4)	1	0.883
Percentage of supply of PPE Kits (X_5)	1	0.865
Supply of Total N95 Masks (X_6)	1	0.822
Numbers of COVID-19 Special Government Hospitals (X_7)	1	0.605
Availability of Government Hospital Beds for COVID Patients (X_8)	1	0.899
Availability of Safe Home Beds for COVID Patients (X_9)	1	0.675
Numbers of vaccination sites (X_{10})	1	0.910
Numbers of total vaccinated population (X_{11})	1	0.840
Population Density (X_{12})	1	0.885

Literacy Rate (X_{13})	1	0.492
Percentage of population below 7 years age (X_{14})	1	0.746
Percentage of population more than 65 years old (X_{15})	1	0.895
Per capita Income (X_{16})	1	0.807
Number of total SC and ST population (X_{17})	1	0.697

Source: Based on the author's calculation
Extraction Method: Principal Component Analysis.

Communalities reflect the amount of variance in each variable that has been accounted for in the components. Estimates of the variation in each variable accounted for by all components or factors are known as initial communalities. If we are doing a correlation analysis, this is always set to 1. Communalities of extraction are estimations of the variation in each variable that is accounted for by components. When the extracted commonalities (c) are high (i.e. $c \geq 0.5$), then we can say that the extracted components accurately describe the variables. You may need to extract another component if any of the communalities are very low (i.e. $c \leq 0.5$) in a primary component extraction. Here in our data, the extracted communalities are high for all the variables that mean the principal components have well described the variables.

After successfully tested the KMO statistic and Bartlett's Sphericity test we have confirmed that the PCA will be the appropriate method for our study. The respective commonalities in the above-mentioned table 4 are also significant to express the variation of the selected variables by the principal components. After running the PCA we have fewer independent components from 17 variables to two components. These two components together can explain the 81% variance in the COVID Vulnerability Index (CVI). Following table 5 have shown the percentage share of variation of each component with their respective Eigenvalues.

Table 5: Percentage of Variation Explained by the Principal Components

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.49	61.68	61.68	10.49	61.68	61.68
2	3.30	19.40	81.08	3.30	19.40	81.08
3	0.87	5.11	86.19			
4	0.67	3.94	90.13			
5	0.62	3.64	93.76			
6	0.37	2.19	95.95			
7	0.27	1.56	97.51			
8	0.15	0.86	98.37			
9	0.10	0.58	98.94			
10	0.06	0.34	99.28			
11	0.05	0.26	99.54			
12	0.04	0.23	99.77			
13	0.02	0.12	99.90			
14	0.01	0.06	99.96			
15	0.01	0.04	100			
16	0.00	0.00	100			
17	0.00	0.00	100			

Source: Based on the author's calculation

The percentage of variation explained by the successive components should be less and less. Here the first component has the highest percentage share of variation 61.68% and the second component is explained as the remaining 19.40% of the total explained variation of 81.08%. And also the first component has the highest Eigenvalue (10.49) followed by the second component (3.30). Hence according to the Kaiser Criteria, we have selected the only first two components and their respective proportion of variation explained to construct the final CVI. A scree plot represents the relationship between the principal components and their respective

Eigenvalues which is shown in the following figure. In addition, it allows us to determine which factors have been able to explain the association between them in a suitable way.

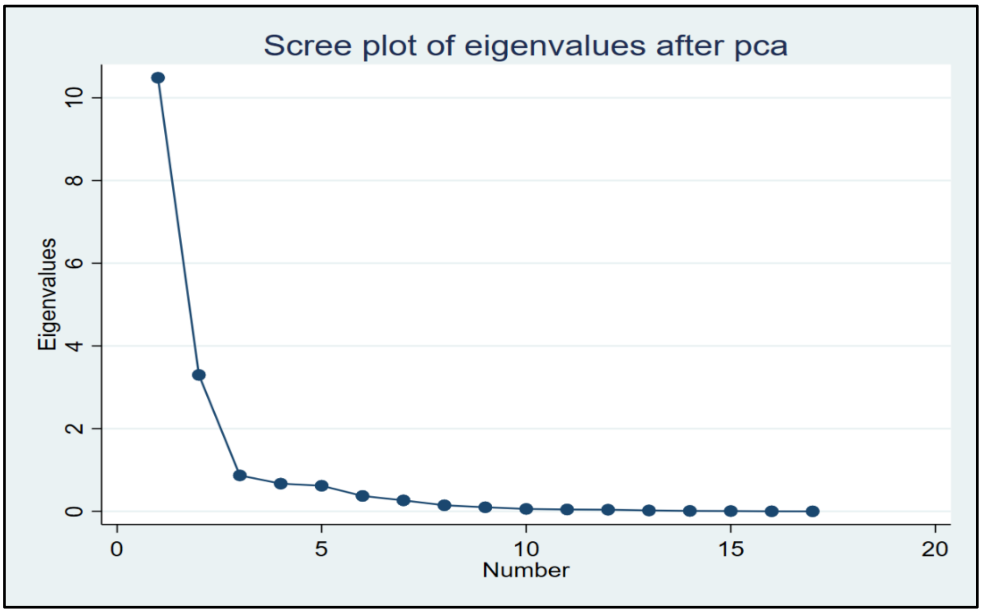


Figure 1: Scree Plot of Principal Components and their respective Eigenvalues

Source: Based on the author's calculation of PCA

After extracting the principal components, we have used orthogonal varimax rotation to the components so that the components are fitted the sample data more adequately. The rotated components coefficient matrix of the selected 17 variables and their respective highest factor loadings have shown in the following table 6. The final CVI scores of West Bengal districts are calculated by adding up the district wise extracted factor loadings multiplied by the percentage of variation explained by the two components respectively.

Table 6: Rotated Component Score Coefficient Matrix

All Factor Loadings		CVI Indicators	Highest Factor Loadings	
PC1	PC2		PC1	PC2
0.923	0.269	Total Positive Cases of COVID-19	0.923	
0.923	0.270	Total Discharged Persons after COVID-19	0.923	
0.940	0.176	Total Deaths from COVID-19	0.940	
0.891	0.299	Total Active Cases	0.891	
0.930	0.019	Percentage of supply of PPE Kits	0.930	
0.905	-0.045	Supply of Total N95 Masks	0.905	
0.716	0.305	Numbers of COVID-19 Special Government Hospitals	0.716	
0.948	-0.028	Availability of Government Hospital Beds for COVID Patients	0.948	
0.075	0.818	Availability of Safe Home Beds for COVID Patients		0.818
0.954	-0.005	Numbers of vaccination sites	0.954	
0.709	0.580	Numbers of total vaccinated population	0.709	
0.882	-0.328	Population Density	0.882	
0.680	0.170	Literacy Rate	0.680	
-0.018	0.863	Percentage of population below 7 years age		0.863

0.469	0.821	Percentage of the population more than 65 years old		0.821
0.884	-0.161	Per capita Income	0.884	
-0.148	0.821	Number of total SC and ST population		0.821

Source: Based on the author's calculation of PCA

As a result, we have used these two rotating principal components and their respective percentage share of explained variance to calculate the final *COVID Vulnerability Index*. The resulting rotated components have been multiplied by their respective % of variance explained. As an outcome, the final CVI scores for each of the West Bengal districts have been calculated. According to the following table 7, we can see that the most vulnerable districts in West Bengal after the COVID-19 pandemic are the Kolkata, North 24-Parganas, Burdwan, Howrah, South 24-Parganas and Hooghly. On the other hand, the less COVID Vulnerable districts are Dakshin Dinajpur, Uttar Dinajpur, Puruliya, Malda, and Cooch Behar. The districts like Bankura, Murshidabad, Darjeeling, etc. are moderately safer as compare to the highly vulnerable zones.

Table 7: Composite Index of COVID Vulnerability in West Bengal Districts

Districts	Factor 1	Factor 2	CVI Score	Ranking
Dakshin Dinajpur	-0.790	-1.020	-68.527	1
Uttar Dinajpur	-0.848	-0.575	-63.440	2
Puruliya	-0.681	-0.794	-57.414	3
Malda	-0.711	-0.391	-51.462	4
Coochbehar	-0.634	-0.436	-47.571	5
Birbhum	-0.417	-0.303	-31.599	6
Bankura	-0.435	-0.208	-30.863	7
Nadia	-0.338	0.118	-18.548	8
Murshidabad	-0.536	0.758	-18.340	9
Darjeeling	0.088	-1.141	-16.723	10
Jalpaiguri	-0.287	0.332	-11.273	11
Purba Medinipur	0.028	-0.149	-1.194	12
Paschim Medinipur	-0.118	0.475	1.935	13
Hooghly	0.179	-0.157	7.995	14
South 24-Parganas	-0.183	1.811	23.841	15
Howrah	0.526	-0.324	26.135	16
Burdwan	0.281	0.954	35.874	17
North 24-Parganas	1.412	2.575	137.037	18
Kolkata	3.464	-1.524	184.136	19

Source: Based on the author's calculation

The top 5 highly COVID vulnerable areas in West Bengal and the comparatively less vulnerable districts till February 2021 are shown in the following table 8 and table 9 respectively.

Table 8: Top 5 COVID Vulnerable zones

Districts	CVI	Ranking
Kolkata	184.136	19
North 24-Parganas	137.037	18
Burdwan	35.874	17
Howrah	26.135	16
South 24-Parganas	23.841	15

Source: Based on the author's calculation

Table 9: Top 5 less COVID Vulnerable zones

Districts	CVI	Ranking
Dakshin Dinajpur	-68.527	1
Uttar Dinajpur	-63.440	2
Puruliya	-57.414	3
Malda	-51.462	4
Cooch Behar	-47.571	5

Source: Based on the author's calculation

The next figure shows a map of COVID vulnerability in West Bengal districts, followed by a measure of spatial dependency using the local Moran's I index.

We have categorized the total range of CVI scores into three parts such as "Highly Vulnerable", "Moderately Vulnerable" and "Less Vulnerable". The description of these categories is discussed in the following table 10.

Table 10: Different Categories of CVI Scores of West Bengal Districts

Categories	CVI Scores	Description
Highly Vulnerable	185.00 – 1.00	The districts are highly vulnerable to the COVID situation. People living in this district are not safe from the life-threateningCOVID-19 pandemic and its recommended to take more actions to prevent the spread of coronavirus.
Moderately Vulnerable	1.00 – -35.00	These areas are comparatively safe to the COVID-19 but also the risk of COVID-19 is not zero inthese regions.
Less Vulnerable	-35.00 – -70.00	Compared to the other locations, the districts in this range are significantly less vulnerable to COVID-19. These districts may not need to take urgent concerns.

Source: Author's classification based on CVI scores

Figure 2: West Bengal districts map

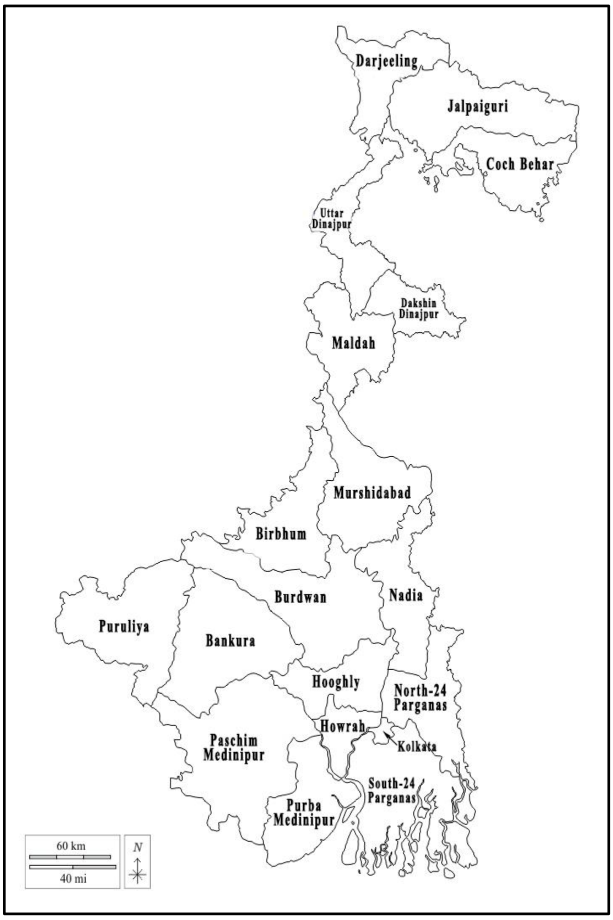
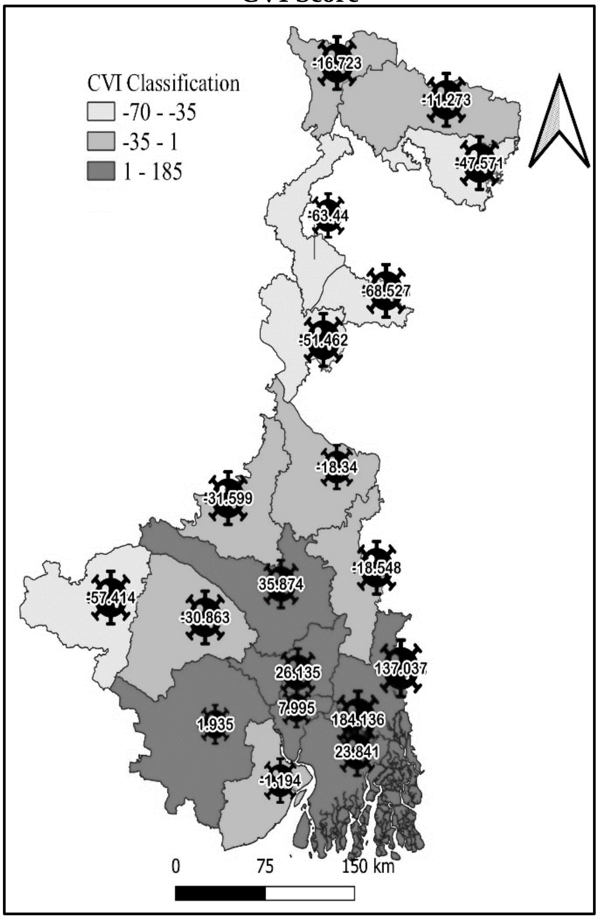


Figure 3:West Bengal districts map according to CVI Score



It can be seen that the districts like Kolkata, South 24-Parganas, North 24-Parganas, Paschim Medinipur, Hooghly, Burdwan are the highly vulnerable regions to COVID-19. Bankura, Purba Medinipur, Nadia, Murshidabad, Jalpaiguri and Darjeeling are comparatively less vulnerable but the risk of spread of COVID is not zero. And the districts of Uttar Dinajpur, Dakshin Dinajpur, Cooch Behar, Maldah and Puruliya have a significantly lower risk of COVID pandemic according to our study.

Using our previously established CVI scores, we have now examined for any spatial autocorrelation among the districts of West Bengal. As we mentioned earlier, we have measured the Local Moran's I correlation coefficient regarding the district respective CVI scores. A positive coefficient or value closer to one represents the existence of spatial dependency among the districts, zero coefficient value implies no spatial correlation and value less than zero or negative coefficient indicates there is a negative spatial correlation. After running the

Moran's I test in the GeoDa[5] software we have got the coefficient score as 0.389, which confirms that there is a spatial correlation existed considering the COVID Vulnerability Index.

Figure 4: Local Moran's I Correlation Coefficient of CVI Scores

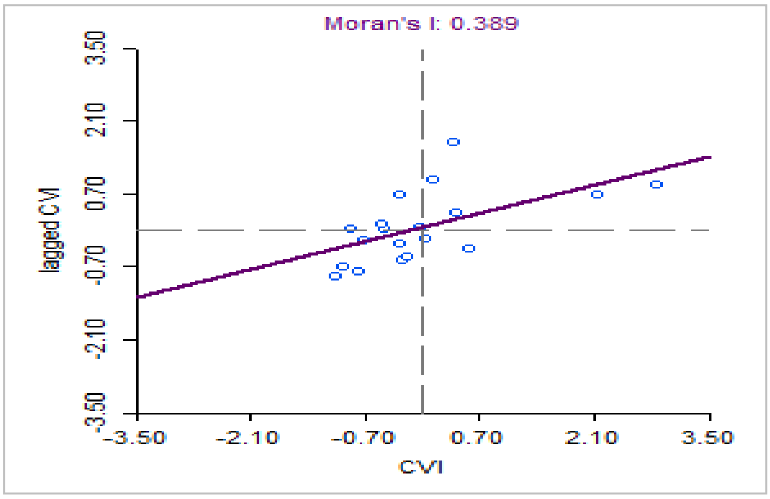


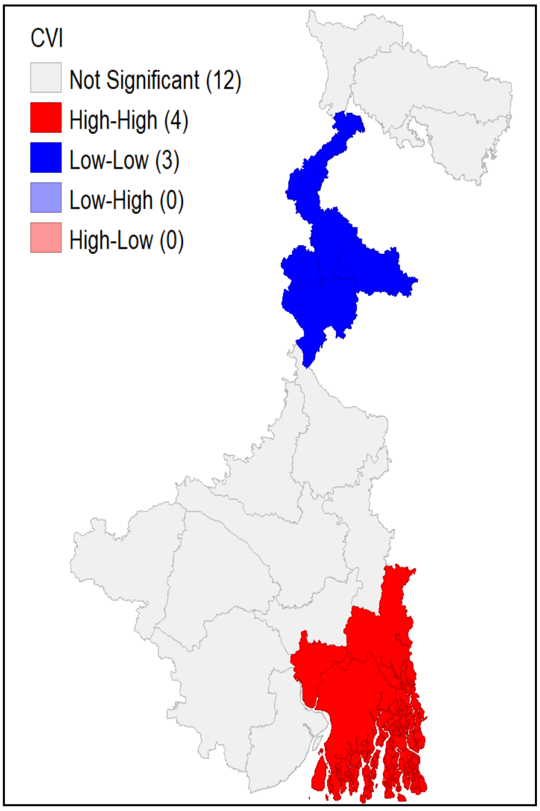
Table 11: Univariate Local Moran's I Estimation

Variable	Moran's I Correlation Coefficient [I]	E[I]	S.D.[I]	z-statistics	Pseudo p-value
CVI	0.389	-0.055	0.156	2.830	0.015*

Source: Based on Author's estimation
**Significant at 5% level of significance*

Hence the positive index value suggesting that the existence of spatial correlation among the districts in terms of their CVI scores and the index score is also statistically significant at a 5% significance level (Pseudo p-value<0.05).Now the significant clustering of "High-High" CVI scores or "Low-Low" CVI scores can be shown by the *Local Indicators of Spatial Association (LISA)* mapping. The "High-High" clusters indicate that the districts with high vulnerability to COVID are adjacent to each other and the "Low-Low" cluster of districts refers to the situation where a less vulnerable district has its neighbouring districts whose COVID vulnerability are also low. The red regions are showing a significant "**High-High**" cluster of the CVI while the blue areas are the representation of the "**Low-Low**" cluster of the same. There is no "High-Low" or "Low-High" clusters of CVI noticed in our study area. So far the "High-High" cluster of vulnerable areas is surrounded in the districts like **Kolkata, North 24-Parganas, South 24-Parganas** and **Howrah**. On the other side districts with low CVI scores are clustered among **Malda, Dakshin Dinajpur,** and **Uttar Dinajpur**.

Figure 1: LISA Cluster Map of CVI



V. CONCLUSION

[5]*Anselin, Luc, IbnuSyabri and Youngihn Kho (2006). GeoDa: An Introduction to Spatial Data Analysis.*

After going through the construction of the *COVID Vulnerability Index(CVI)* of West Bengal districts using the secondary data of novel coronavirus pandemic and socio-economic variables till the time February 2021 we have seen that the 7 districts *(Paschim Medinipur, Howrah, Hooghly, Kolkata, North 24-Parganas, South 24-Parganas, Burdwan)*are highly life-threatening to live for coronavirus, again 7 districts*(Birbhum, Bankura, Nadia, Murshidabad, Darjeeling, Jalpaiguri, Purba Medinipur)*are moderately secure to COVID and remaining 5 districts *(Dakshin Dinajpur, Uttar Dinajpur, Puruliya, Malda, Cooch Behar)* are comparatively safe to live. West Bengal districts such as Kolkata, North 24 Parganas, South 24 Parganas, and Howrah are known as developed regions because their per capita income, literacy rate, healthcare facilities, and other socio-economic parameters are much higher than the other districts, but it is clear that these districts also face the threats of higher COVID vulnerability. It might possible because of the higher population and high population density. So we can conclude that the developed regions are needs to take higher concerns on the spread of the virus. Declaring several lockdowns by the Central and State Government after the first and second waves of the novel coronavirus, closing the private service sectors by ordering the workers to work from home, working the government service sectors giving access to limited persons, terminating all educational institutions until the further noticed, closing the supermarkets, cinema halls and any places involved crowding, providing free daily meals to the backward section of the society through rationing system,increasing the awareness through different social communications, providing free sanitizer and one-time useable masks, and so many strategies have been taken by the government to prevent the spread of the disease.

REFERENCES

[1]. Age Poom, O. J. (July-December, 2020). COVID-19 is spatial: Ensuring that mobile Big Data is used for social good. Big Data & Society, journals.sagepub.com/home/bds, DOI: 10.1177/2053951720952088, 1-7.

[2]. Apurba Sarkar, P. C. (2021). COVID-19: District level vulnerability assessment in India. Elsevier, Clinical Epidemiology and Global Health, , 204-2015.

[3]. D.J. Abson, A. J. (2012). Using principal component analysis for information-rich socio-ecological vulnerability mapping in Southern Africa. Appl. Geogr 35, 515-524.

[4]. Emmanuel Mavhura, B. M. (2017). An approach for measuring social vulnerability in context: The case of flood hazards in Muzarabani district, Zimbabwe. Elsevier, Geoforum 86, 103-117.

[5]. Fernando Flores Tavares, G. B. (2020). The Pandemic of Poverty, Vulnerability, and COVID-19: Evidence from a Fuzzy Multidimensional Analysis of Deprivations in Brazil. World Development, https://doi.org/10.1016/j.worlddev.2020.105307.

[6]. HS. Gopalan, A. M. (2020). COVID-19 pandemic and challenges for socio-economic issues, healthcare and National Health Programs in India. Diabetes Metab Syndr Clin Res Rev. 14(5), 757-759.

[7]. I.M. Karaye, J. H. (2020). The impact of social vulnerability on COVID-19 in the U.S.: an analysis of spatially varying relationships. Am J Prev Med. 59(3). https://doi.org/10.1016/j.amepre.2020.06.006, 317-325.

[8]. K. Chatterjee, A. K. (2020). Healthcare impact of COVID-19 epidemic in India: a stochastic mathematical model. Med J Armed Forces India. 76(2) https://doi.org/10.1016/j.mjafi.2020.03.022, 147-155.

[9]. Li Zhenlong, L. X. (2020). Monitoring the Spatial Spread of COVID-19 and Effectiveness of Control Measures Through Human Movement Data: Proposal for a Predictive Model Using Big Data Analytics. JMIR Research Protocol, Vol. 9, Issue 12, 1.

[10]. Patricia Lustosa Brito, M. K. (September, 2020). The Spatial Dimension of COVID-19: The Potential of Earth Observation Data in Support of Slum Communities with Evidence from Brazil. ISPRS International Journal of GEO-INFORMATION.

[11]. Q Li, X. G. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med, 382(13), 1199-1207.

[12]. R. Acharya, A. P. (2020). A vulnerability index for the management of and response to the COVID-19 epidemic in India: an ecological study. Lancet Glob Heal, https://doi.org/10.1016/S2214-109X(20)30300-4.

[13]. Rodrigo Barraza, R. B. (May, 2020). COVID-19 and vulnerability: a multidimensional poverty perspective in El Salvador. UNDP Latin America and the Caribbean, #COVID19 | POLICY DOCUMENTS SERIES.

[14]. S. Gupta, G. R. (2020). Effect of weather on COVID-19 spread in the US: a prediction model for India in 2020. Sci Total Environ, 728:138860. https://doi.org/10.1016/j.scitotenv.2020.138860.

[15]. Shrikrushna Subhash Unhale, Q. B. (2020). A Review of Corona Virus (COVID-19). World Journal of Pharmaceutical and Life Science, Vol. 6, Isuue 4. ISSN 2454-2229, 109-115.

[16]. Y.S. Chao, C. W. (2017). Principal component-based weighted indices and a framework to evaluate indices: results from the Medical Expenditure Panel Survey 1996 to 2011. PloS One. 12(9). https://doi.org/10.1371/journal.pone.0183997.

Souvik Dey. "Measuring the Risk of COVID-19 Pandemic in West Bengal Districts." *International Journal of Humanities and Social Science Invention (IJHSSI),* vol. 11(10), 2022, pp 90-102. Journal DOI- 10.35629/7722