RESEARCH ARTICLE

A Prediction of Spatial Distribution for Epidemic Infectious Diseases Using a Proposed Deep Learning Framework Based on GIS (Case Study: COVID-19 in Egypt)

Ismail S. Tawfik *, Prof. Dr. Christina Albert **

*(PhD Candidate, Computer and Information Systems Department, Sadat Academy for Management Sciences, Cairo Egypt, Email: Ismail.tawfik@gmail.com)

** (Prof. of Computer and Information Systems Department, Sadat Academy for Management Sciences, Cairo Egypt Email : sams.christina.albert@gmail.com)

----------**********----------

Abstract:

COVID-19 has impacted over 170 countries worldwide, causing a significant increase in infections and deaths. On March 11, 2020, the World Health Organization (WHO) officially declared COVID-19 a pandemic. In response, several strategies based on spatial and deep learning approaches, were employed to predict high-risk areas. This study proposes a deep learning framework based on Geographic Information Systems (GIS) to predict the spatial distribution of COVID-19 risk in Egypt. Four key indicators demographic, residential, environmental, and atmospheric were developed using geomatics technology, in alignment with the United Nations' Sustainable Development Goals (SDGs), specifically Targets 11 and 3. The integration of deep learning and GIS identified high-risk governorates such as Cairo, Giza, Alexandria. These regions are characterized by factors such as overcrowding, high population density, inadequate infrastructure, and environmental conditions like air pollution and seasonal weather changes, all of which contribute to the spread of the virus. The framework predictions closely aligned with official data from the Egyptian Ministry of Health and Population (MOHP). This research provides a valuable framework for identifying high-risk areas and supports the health sector in developing strategies to mitigate the spread of the pandemic. The proposed framework offers actionable insights for decision-makers worldwide in their efforts to control future outbreaks.

Keywords — **World Health Organization (WHO), Ministry of Health and Population (MOHP), Rectified Linear Unit (ReLU), Mean Squared Error (MSE).**

----------**********----------

I. INTRODUCTION

COVID-19, caused by the SARS-CoV-2 virus, was first reported in late 2019 in Wuhan, Hubei Province, China (WHO,2020,82). By early 2020, the virus spread rapidly, leading to global concern as it affected countries across all continents. On March 11, 2020, the World Health Organization (WHO) declared COVID-19 a pandemic, which has since affected approximately 215 countries and territories worldwide [1]. The rate of infection and mortality varied significantly among different regions [2], with over 83 million confirmed cases and 1.84 million deaths globally by January 4, 2021 [1]. COVID-19 primarily causes acute respiratory illness, with a global mortality rate estimated at 1.38% [1][3].

In Egypt, the virus has had a significant impact, with 142,187 confirmed cases and 7,805 deaths as of the latest update, and these numbers continue to rise daily [1]. COVID-19 is primarily transmitted

via droplets or aerosols expelled when an infected person coughs or sneezes [4][5]. Symptoms typically appear within 2 to 14 days, although some individuals may remain asymptomatic [4][5]. The rapid spread of the virus has posed substantial health and logistical challenges for Egypt.

This study aims to implement a proposed framework that integrates deep learning techniques, such as deep neural networks, with Geographic Information Systems (GIS) techniques, including spatial analysis, by using tools like MATLAB and ArcGIS. The framework will be used to analyse the spread patterns of COVID-19, with a particular focus on identifying high-risk areas. The ultimate goal is to support public health decision-making by providing valuable insights that can guide targeted interventions and resource allocation for regions most vulnerable to COVID-19.

II. STUDY AREA

This study was conducted in Egypt, located in North Africa, with a population of approximately 105 million people (CAPMAS, 2017) and covering an area of 1,002,000 km². Egypt is administratively divided into 27 governorates, each with varying levels of population density, infrastructure, air pollutions, weather seasonal changes and access to healthcare services. These differences make the governorates crucial for the analysis of factors affecting the spread of infectious diseases. These governorates are important to the study as they differ significantly in population density, healthcare access, and infrastructure etc. which could impact the spread of epidemic diseases. Fig. 1 provides an illustration of the geographical layout of Egypt's governorates. According to statistics from the Ministry of Health and Population (MOHP), certain Egyptian governorates, including Cairo, Alexandria, and Giza, have seen some of the highest rates of COVID-19 transmission. These regions were prioritized in this study due to their high population density and the significant number of confirmed COVID-19 cases, which make them hotspots for the virus's spread. The focus of this study aims to investigate the factors contributing to the spread of the virus in these regions and to develop a framework for predicting which governorates are most vulnerable to future outbreaks of the pandemic.

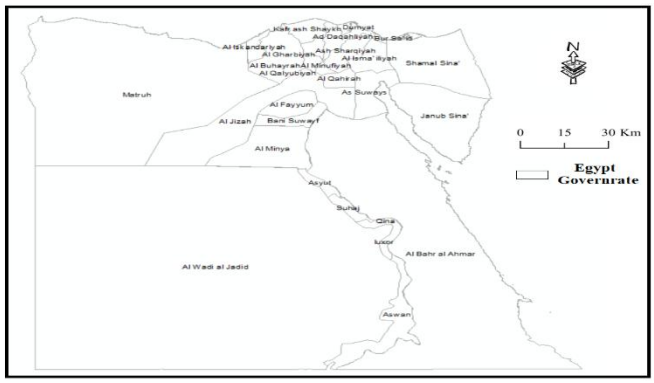

Fig. 1, Location of Study Area

III. IMPACT OF THE SELECTED NDECATORS ON COVID-19 TRANSMETION

According to [6][7][8], The spread and severity of COVID-19 are influenced by a complex set of factors, including residential, demographic, environmental, and weather indicators. These factors can significantly affect the vulnerability and risk of COVID-19 outbreaks in different regions. Below, the researcher explains how each of these indicators are associated with COVID-19 risk areas, supported by references from relevant studies see fig. 2.

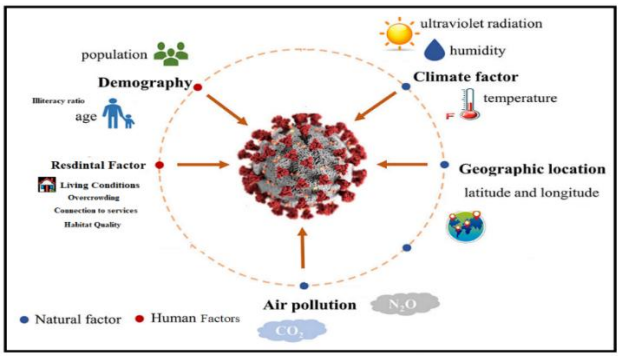

Fig. 2 Influence of the selected indicators and the risk areas for COVID-19 transmission, by Researcher

A. Residential Indicators:

Residential factors, such as housing density, crowding, and living conditions, have been linked to the increased spread of COVID-19.

1) ***Living Conditions:*** Poor housing conditions, such as lack of access to clean water and sanitation, can exacerbate the spread of infectious diseases. Studies have found residential areas with poor infrastructure (e.g., inadequate sewage systems or overcrowded housing) have higher transmission

rates of COVID-19. In Barcelona, individuals living in overcrowded conditions or with damp housing experienced significantly higher COVID-19 prevalence, indicating a direct link between housing quality and disease spread [9]. These conditions hinder effective preventive measures, such as social distancing and frequent hand washing.

B. Demographic Indicators

Demographic factors such as population, age, illiteracy ratio and habitat quality have a well-established influence on COVID-19 spread.

1) ***Population Density:*** As stated in [10], the study focused on high population density and its correlation to higher virus spread and intensity which is one of the demographics factors. Regions with higher population density tend to have greater challenges in controlling the spread of infectious diseases. COVID-19 transmission is significantly higher in densely populated areas because of the increased likelihood of close contact between individuals. Crowded residential areas facilitate the spread of respiratory droplets, a primary mode of transmission for COVID-19.

2) ***Age*****:** The impact of age on COVID-19 is multifaceted, influencing both health outcomes and societal responses. Older adults are particularly vulnerable to severe illness and hospitalization due to age-related immune decline and comorbidities. This demographic has faced unique challenges during the pandemic, including restrictive regulations that have affected their mobility and social interactions. The Older adults, especially those aged 85 and above, exhibit the highest hospitalization rates, with incidence rates reaching up to 5889.6 per 100,000 persons [11]. following sections outline key aspects of how age affects COVID-19 outcomes.

C. Environmental Indicators:

Environmental factors such as air pollution, urbanization, and pollutants have been linked to the increased severity of COVID-19.

1) Air Pollution: The impact of air pollution on COVID-19 has been a significant area of research, revealing complex interactions between environmental factors and viral transmission. Studies indicate that air pollutants, particularly particulate matter (PM2.5 and PM10), nitrogen dioxide (NO2), and ozone (O3), not only influence the severity of COVID-19 but also affect mortality rates associated with the disease. The following sections elaborate on these findings.

Increased levels of PM2.5 and NO2 have been linked to higher COVID-19 morbidity and mortality rates, as these pollutants can enhance susceptibility to the virus by affecting immune responses [12].

While the reduction of air pollution during the pandemic has shown potential health benefits, the long-term implications of air quality on respiratory diseases and viral infections remain a critical concern for public health policy.

D. Weather Indicators

Weather conditions, including temperature, humidity, and seasonality, have been shown to influence the transmission and spread of COVID-19.

1) ***Temperature and Humidity:*** As stated in (Wang et al., 2020) in his study found, under a linear regression framework, high temperature and high humidity significantly reduce the transmission of COVID-19, respectively. One-degree Celsius increase in temperature and one percent increase in relative humidity lower R by 0.0225 and 0.0158, respectively. result is consistent with the fact that the high temperature and high humidity reduce the transmission of influenza and SARS.

2) ***Seasonality:*** There are indications that seasonal variations (such as the shift from winter to spring) may influence the intensity of outbreaks. The virus's survival rate may decrease in warmer months, though COVID-19 transmission continues to occur in all seasons, making it clear that other factors, such as human behaviour and public health measures, play a much larger role in transmission.

The correlation between residential, demographic, environmental, and climate indicators and COVID-19 risk areas is well-documented. High population density, inadequate housing and air pollution all contribute to an increased risk of COVID-19 transmission and severity. Additionally, climate factors may influence the virus's spread, though their impact is less significant compared to other indicators such as human behaviour and healthcare infrastructure. Understanding these correlations helps in predicting high-risk areas and developing targeted interventions for controlling the pandemic.

IV. DATA COLLECTIONS AND METHODOLOGY

A. Data

The data used in this study were obtained from different sources Table 1. The Demographic and Residential data were collected from the official reports of the central agency for public mobilization & statistics (CAPMAS). Climate data were obtained from Wunderground website, Remote sensing data included Sentinel-5 for NO2 and CO pollutant gases.

Many previous studies have explored the impact of demographic, residential, environmental and weather factors on the spread of infectious diseases in general and COVID-19 in particular.

1) ***Demographic Data:*** The relationship between population demographics – including density – and COVID-19 estimates are important for the planning of policy and medical resources. Population density has been used as a surrogate measure of social distancing capacity and studies

have shown that SARS-CoV-2 transmission is potentially more likely to occur among cities with higher population densities. Also increased age has been reported to be a factor for COVID-19 severe outcomes. literacy has a positive and significant contribution to the recovery rate out of the total individual tested as well as the total vaccination [13].

The researcher has selected the most demographic sub-factors affect the spread of COVID-19 such as population density, Population Per (km2), Illiteracy ratio, old age ratio to apply the proposed framework and data was obtained from CAPMAS annual reports.

2) ***Residential Data:*** A poor connection to services and habitat quality has extremely contributed to increase the spread of COVID-19 in many countries specially the Third World countries. Also overcrowding plays a major role in the spread of the COVID-19. The combined risk of overcrowding on both underlying health conditions and the spread of infectious diseases may mean that areas with high rates of household overcrowding also experience higher rates of COVID-19 transmission and mortality [14].

The researcher selected key residential sub factors that influence the spread of COVID-19, including sanitation networks, electricity networks, households with a single room, households with access to water and soap, and common toilet facilities, to apply within the proposed framework. data was obtained from CAPMAS annual reports for the study regions.

As stated in (Haque & Rahman,2020), study have suggested that under a linear regression framework, high temperature and high humidity significantly reduce the transmission of COVID-19, respectively.

3) ***Pollution Data***: NO2 and CO pollutant gases were studied to investigate the effect of air pollution as an indicator for COVID-19 spreading prediction. NO2 and Co are considered the most dangerous and harmful air pollutants in urban areas [15].

4) ***Weather Data:*** Previous studies have demonstrated that meteorological factors, particularly temperature and absolute humidity can significantly affect the transmission of respiratory infectious diseases, including SARS, influenza, and tuberculosis [16]. There was a negative relationship between COVID-19 cases per million and temperature to the number of days since the first case was reported [17].

The researcher selected maximum and minimum temperature, along with absolute humidity data, as key factors influencing the spread of COVID-19 to apply in the proposed framework. Historical weather data was obtained from Wunderground.com for the year of 2020,2021,2022 for 36 Months for the study regions. Table 1

TABLE 1
DATA SOURCE

Criteria	Description	Data Extraction time	Type	Source
P	Population Density	2017	Vector	(CAMPAS)
PD	Population Per (km)	2017		
IR	Illiteracy Ratio	2017		
OA	Old Age Ratio	2017		
SN	Sanitation Networks	2017		
EC	Electricity Connections	2017		
WS	Households with Water and Soap	2014		
CT	Common toilet	2017		
CR	Households Crowding Ratio	2017		
PO	Pollution	10/2020	Raster	Sentinel-5 Satellite image with high resolution
XT	Maximum Temp.	01-01-2020 To 31/12/2022	Vector	Wundergrou nd.com
MT	Minimum Temp.		Vector	
HU	Humidity		Vector	

B. Methodology

A proposed deep learning framework integrated with Geographic Information Systems (GIS) was developed to predict the spatial distribution of COVID-19 risk areas. This approach is based on previous studies and aligns with the guidelines of the UN-Habitat Sustainable Development Goals (SDG 3 & 11), which focus on ensuring safe, resilient, and sustainable cities. SDG 11 aims to make cities and human settlements inclusive, accessible, and resilient, while SDG 3 targets health and well-being for all.

The methodology incorporates four key factors: demographic, residential, environmental, and weather. A total of 13 factors were carefully chosen for integration based on their relevance to COVID-19 vulnerability. These key factors include variables such as population density, illiteracy rates, access to essential services, air pollution levels, and climate conditions. A detailed overview of these factors is presented in Table 2.

COVID-19 spatial prediction scenarios based on the demographic, residential, and environmental factors

relevant to the spread of COVID 19 were developed using GIS data, while COVID-19 spatial prediction scenario based on the weather factor was developed using deep learning techniques. The four prediction scenarios were then integrated to produce a comprehensive prediction scenario of COVID-19 risk areas. The prediction results were compared with official data from the Ministry of Health (MOH) to validate the framework accuracy.

TABLE 2
SELECTED FACTORS FOR SPATIAL PREDICTION OF COVID-19 OUTBREAKS

GIS Factors			
Demographic Indicator			
Sub Factors	**Urban population** Population Density Population Per km2	**Illiteracy ratio**	**Quality of society** old age ratio
Residential Indicator			
Sub Fators	**Overcrowding** Households with one room	**Connection to services** Sanitation networks Electricity networks	**Habitat Quality** Households with Water & Soap Common Toilet
Environmental Indicator			
Sub Factors	**Pollution:** NO2 CO		
Deep learning Factors			
Weather Indicator			
Sub Factors	**Weather Data** Maximum Temperature Minimum Temperature Absolute Humidity		

1) ***Diagram for applied methodology:*** The following diagram explains the applied methodology Fig. 3

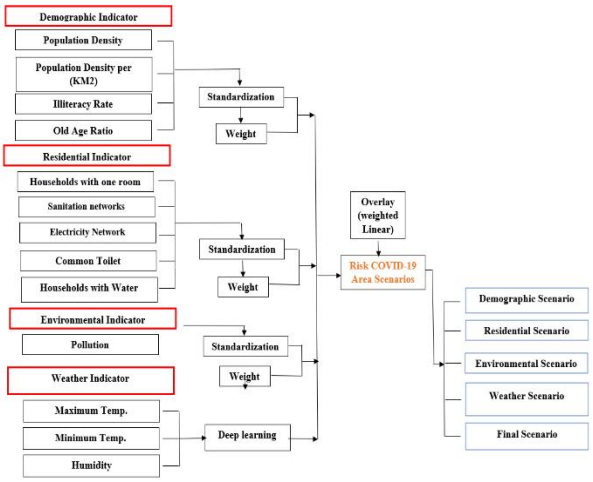

Fig. 3, Diagram for applied methodology, by researcher

2) ***A proposed framework:*** The following Fig. 4 illustrate the proposed framework

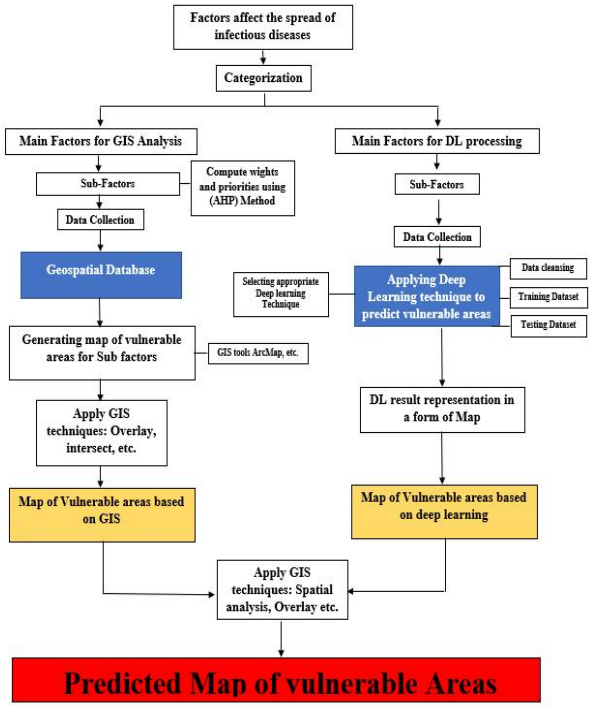

Fig. 4, A Proposed Deep Learning Framework Based on GIS to Predict the Spatial Distribution of Epidemic Infectious Diseases.

V. METHODS

A. AHP method to compute weights and priorities

The researcher selected AHP method to compute weights and priorities. The Analytical Hierarchy Process (AHP) is a method was developed by Thomas L. Saaty in the 1970s for organizing and analysing complex decisions. AHP provides a rational framework for a needed decision by quantifying its criteria and alternative options, and

for relating those elements to the overall goal, in which many variables or criteria are considered.

In the prioritization and selection (quantification) of alternatives [18]. The AHP method uses hierarchical structures to represent a problem and then develop priority scales through the numeric scale calibration for measuring the qualitative performances [18].

The first step in the AHP process is breaking down the problem into smaller, more manageable components. In this research, a set of criteria was developed through an extensive literature review and expert input. These criteria were then categorized into four main categories: demographic, residential, environmental, and weather factors, all of which are relevant to the spread of COVID-19. Each main category consists of a set of sub-factors that contribute to the overall analysis [19][20].

To assign weights for the different sub-factors, the AHP pairwise comparison matrix approach was made to estimate the relative weights using Saaty's scale. see table 5, Table 6. Subsequently, these assigned weights were tested for consistency by computing a consistency ratio (CR), Finally, a square matrix a consistency ratio was checked as the required threshold value should be less than 0.1 by the following equations:

$$CR = \frac{CI}{RI} \qquad CI = \frac{\lambda - n}{n - 1}$$

where CI is the consistency index number of criteria being compared (for a reciprocal matrix, yy max >= n). RI is a random consistency index that depends on the number of criteria, λ is the maximum eigenvector of the matrix, and n is the number of criteria. Where IR is the average value of CIR values for random matrices using the Saaty scale, Table 3. A Saaty's scale was utilized in the pairwise comparison matrix for demographic, residential and environmental factors respectively.

TABLE (1-3)
SAATY'S SCALE

Size of Metrix	1	2	3	4	5	6	7	8	9	10
IR	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

The allowable value of CR should not exceed 0.1 [21]. In this study, a CR value of 0.09 was obtained. This indicates that the sub indicators matrix is consistent.

Each criterion used was classified into Four value classes, each of them was assigned a score from 1 (less important) to 4 (high important) (see Table 7, Table 8) . The selected sub indicators were rasterized and reclassified using different GIS (ArcMap 10.1) tools See Fig. 3, Fig. 4 , Fig. 5.

TABLE 4
FUNDAMENTAL SCALE OF ABSOLUTE NUMBERS RANGING FROM 1 TO 9.

Intensity of Importance	Importance
1	Equal Importance
3	Moderate Importance
5	Strong Importance
7	Very Strong Importance
9	Extreme Importance
2,4,6,8	Values bet ween tow judgment
Reciprocals	Used for invers comparison

TABLE 5
AHP CRITERIA PAIRWISE COMPARISON MATRIX FOR THE CVI OF THE DEMOGRAPHIC INDICATOR

	Population Density	Population Per (km2)	Illiteracy ratio (%)	Old age ratio (%)	Weight
Population Density	1.00	1.00	0.33	3.00	23.75
Population Per (km2)	1.00	1.00	0.33	0.33	13.28
Illiteracy ratio	3.00	3.00	1.00	2.00	43.79
Old age ratio	0.33	3.00	0.50	1.00	20.75

TABLE 6
AHP CRITERIA PAIRWISE COMPARISON MATRIX FOR THE CVI OF THE RESIDENTIAL INDICATOR

Matrix	House holds with one room	No Sanitation networks	No Electricity networks	House holds with no water and soap	House holds Comm on Toilet	Weight
Households with one room	1.00	4.00	3.00	2.00	1.00	33.65
No Sanitation networks	0.25	1.00	3.00	3.00	3.00	28.63

No Electricity networks	0.33	0.33	1.00	0.33	3.00	11.83
Households with no water and soap	0.50	0.33	3.00	1.00	3.00	20.45
Households Common Toilet	1.00	0.33	0.33	0.33	1.00	6.79

VI. IMPLEMENTATION OF GIS USING GEOMATCS TECHNIQUES

Studied indicators were grouped according to proper weightings. Each indicator was produced as a map layer for further analysis. A Standardizing process step was generated to transform the attributes of each indicator into a common suitability index, Tables 7,8,9. These indices reflect a scale number from 1 to 4. The scale 1 refers to the least risk areas for COVID-19 pandemic distribution in Egypt while 4 refers to the most risk areas. Then the weighted overlay process was applied to produce demographic, residential and environmental indicators maps.

TABLE 7
STANDARDIZATION OF THE DEMOGRAPHIC FACTORS.

Scale	Demographic Data			
	Population density	Population density (km2)	Illiteracy ratio (%)	old age ratio (%)
1	< 500000	465.34–17103.02	3.49–14.14	3.77–8.57
2	500000 - 1500000	17103.02–33740.70	14.14–24.78	8.57–13.38
3	1500000 - 3000000	33740.70–50378.38	24.78–35.43	13.38–18.18
4	>3000000	>50378.38	>35.43	>22.99

TABLE 8
STANDARDIZATION OF THE RESIDENTIAL INDICATOR FACTORS.

Scale	Residential Data				
	Households with one room Ratio (%)	Sanitation networks Ratio (%)	Electricity networks Ratio (%)	Households with Water and Soap Ratio (%)	Common Toilet Ratio (%)
1	< 5	0.16–4.35	0.13–1.06	> 98	0.058–6.31
2	5-10	4.35–8.55	1.06–4.09	95-98	6.31–12.56
3	10-20	8.55–12.75	4.09–8.56	90 - 95	12.56–18.81
4	>20	>12.75	>8.56	< 90	> 18.81

TABLE 9
STANDARDIZATION OF THE ENVIRONMENTAL INDICATOR FACTORS.

Scale	Environmental Data	
	NO2 (Mol/m2)	Co*10-4 (Mol/m2)
1	0.000009–0.000014	0.00049–0.00058
2	0.00014–0.000019	0.00058–0.00068
3	0.000019–0.000023	0.00068–0.00078
4	>0.000023	>0.00078

VII. APPLYING GIS TECHNIQUES

Based on the previous approach and depending on the selected sub-indicators, the COVID-19 vulnerability map for Egypt Governorates was developed. The developed map was classified Egypt governorates into four different COVID-19 vulnerability classes (1 = very low, 2 = low, 3 = high, and 4 = very high).

A. STANDARDIZEF MAP FOR DEMOGRAPHIC FACTORS

The researcher utilized GIS tools to create a vulnerability map for COVID-19 distribution in Egypt, based on the standardization of demographic factors. These factors include population density, population density per square kilometre, illiteracy rate, and the elderly population rate, as outlined in Table 7

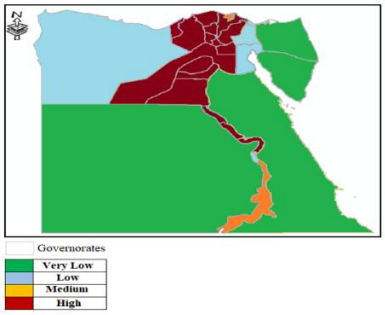

Fig. 5 Scoring for Population density

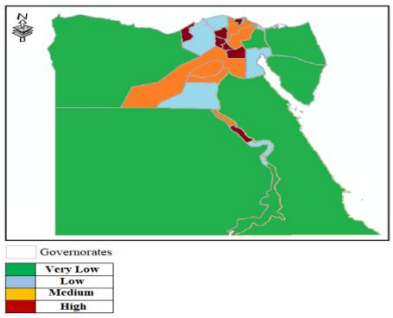

Fig. 6 Scoring for Population density per (km2)

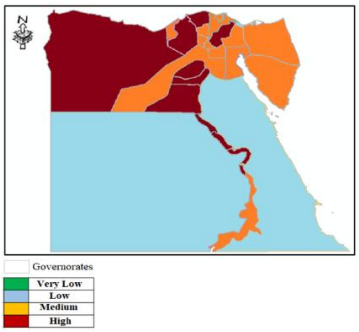

Fig. 7, Scoring for Illiteracy rate

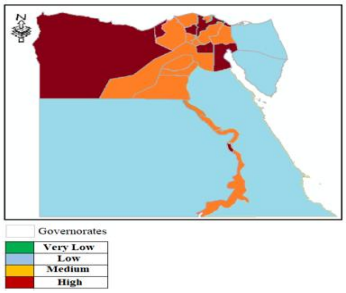

Fig. 8, Scoring for old age rate

B. Standardized map for residential factors

The researcher utilized GIS tools to create a vulnerability map for COVID-19 distribution in Egypt, based on the standardization of residential factors. These factors include electricity network connection, sanitation network connection, households with a common toilet, households with one room, and households with water and soap, as outlined in Table 8.

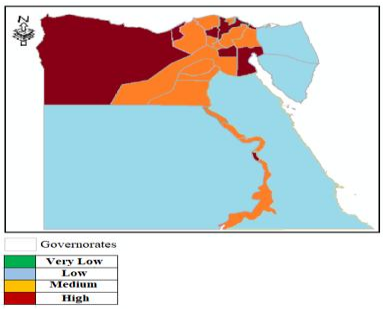

Fig. 9, Scoring for household with electricity network connection

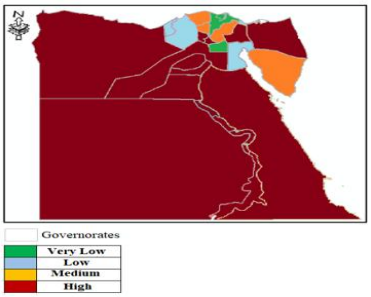

Fig.10, Scoring for household with Sanitation network connection

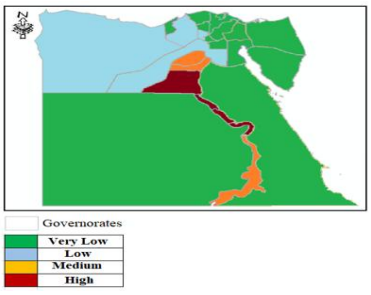

Fig. 11, Scoring for household with common toilet

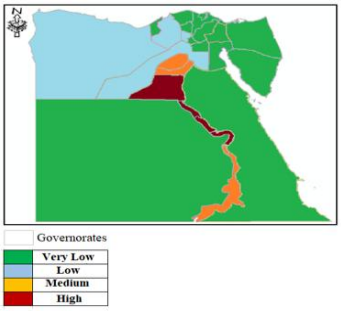

Fig. 12, Scoring for household with one room

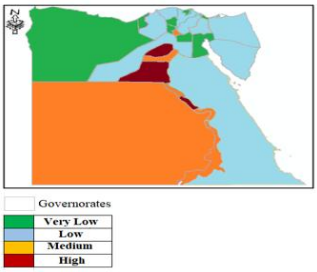

Fig. 13, Scoring for household with water and soap

C. Standardized Map for Environmental factors

The researcher utilized GIS tools to create a vulnerability map for COVID-19 distribution in Egypt, based on the standardization of environmental indicator factors, including NO2 and CO levels Table 9.

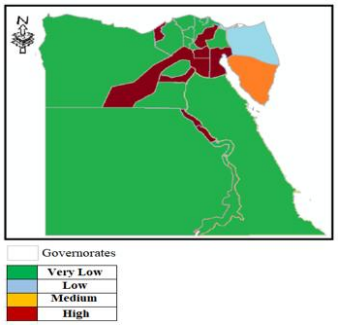

Fig. 14, Scoring for No2

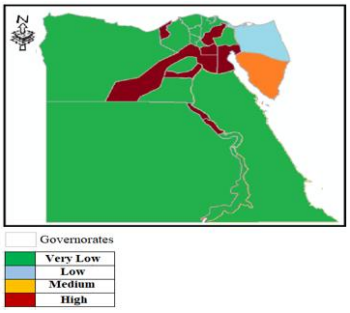

Fig. 15, Scoring for CO

D. Spatial distribution map for COVID-19 based on GIS

Based on the methodology outlined earlier and using the selected demographic, residential, and environmental sub-factors for GIS analysis, a COVID-19 vulnerability map for Egypt was created (see Fig. 16). The map classifies the Egyptian governorates into four vulnerability categories: very low, low, medium, and high, with equal weights assigned to each sub-factor. The analysis was conducted using data from Tables 7, 8, and 9, along with figures Fig.5, Fig. 6, Fig. 7, Fig. 8, Fig. 9, Fig. 10, Fig. 11, Fig. 12, Fig. 13, Fig. 14 and Fig.15, employing the overlay-intersect technique in GIS ArcMap.

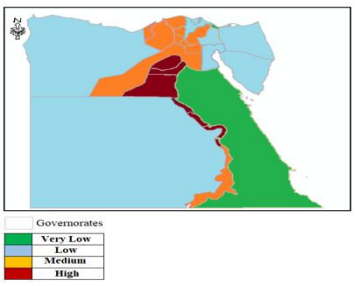

Fig.16, Scoring for all factors scenarios (demographic, residential and environmental factors).

VIII. Results of Applying GIS Techniques

The application of the overlay technique in GIS to assess COVID-19 vulnerability across Egyptian governorates yielded valuable insights into the spatial distribution of risk areas Fig. 16. By combining demographic, residential, and environmental factors, a comprehensive vulnerability map was created, which classified the governorates into four risk categories: very low, low, medium, and high.

A. Demographic Factors Analysis

The demographic layer, which included factors such as population density and the proportion of elderly residents, played a crucial role in identifying regions at higher risk of COVID-19 transmission. For instance, densely populated governorates such as Cairo and Giza were categorized as "high" and "very high" risk, highlighting the vulnerability of these areas due to overcrowding, which facilitates the spread of infectious diseases. On the other hand, less populated areas, like El Wadi El Gidid and Matrouh, were classified as "low" risk due to their lower population density.

B. Residential Factors Analysis

Residential factors, including sanitation and electricity network connectivity, were also integrated into the analysis. Governorates with poorer access to essential services, such as sanitation and water supply (e.g., rural areas), exhibited higher vulnerability, as these conditions increase the potential for virus transmission. For example, governorates like Qena and Sohag were marked as "high" risk due to a lack of adequate infrastructure, contributing to the spread of the virus. In contrast, urban centres with more reliable services, such as Alexandria, were ranked with a "medium" risk due to better infrastructure but still faced high population density challenges.

C. Environmental Factors Analysis

Environmental factors, such as air pollution (NO2, CO levels), were also integrated into the overlay analysis. Areas with high pollution levels, such as Cairo, exhibited an increased risk due to the synergistic effects of poor air quality on respiratory health, which can exacerbate the severity of COVID-19 symptoms. This resulted in an even higher vulnerability classification for those areas. Conversely, areas with lower pollution, such as South Sinai, were classified with lower risk levels, as the environmental conditions were more favourable to public health.

D. Final Vulnerability Map resulting from applying GIS techniques

The final vulnerability map (Figure 1-16) reflects the combined influence of all three factors, providing a comprehensive view of COVID-19 risk in Egypt. The overlay-intersect technique allowed for the identification of high-risk governorates, such as Cairo, Giza, and Luxor, while areas like Matrouh and El Wadi El Gidid were categorized in the "very low" vulnerability class. These results are aligned with the known patterns of COVID-19 transmission in Egypt, validating the effectiveness of the GIS overlay technique in identifying areas most vulnerable to the pandemic.

IX. Implementation of the Deep Learning Techniques Using Weather Data to Predict Spatial Distribution for COVID 19

This part of research concerns the practical implementation of the deep learning techniques for the proposed framework for predicting the spatial distribution of epidemic infectious diseases. It begins with a description of the dataset used, followed by the pre-processing steps, design, training, evaluation, and results.

A. Dataset Collection and Pre-processing

A critical component of deep learning is the quality and preparation of the data. This section explains how the dataset was collected, the features included, and the pre-processing steps taken to prepare the data for training and testing the deep learning model.

1) ***Data Collection:*** The dataset used for the implementation of deep learning techniques consists of daily weather data (Maximum temperature, Minimum temperature and absolute humidity) which affecting the spread of COVID -19. The data was collected for the period of [1095 days] for the years [2020,2021,2022] for the region of [Egypt governorates] with 30000 samples and 3 features. The researcher extracted and explained the 3 features in the following section:

Extraction of temperature and relative humidity values that affect the spread of the COVID-19 from previous studies:

The researcher summarized the previous studies in the following points:

A. Temperature Range

- •Above 11°C: Virus transmission is hindered (Kassem, 2020).
- •5°C−14°C: Optimal temperature range for COVID-19 spread, with a peak at 10°C (Kassem, 2020).
- •Every 1°C increase in temperature decreases cumulative cases by 0.86 (Kassem, 2020).
- •13-24°C: Most suitable temperature range for virus activity and transmission (Anis, 2020).
- •Cities with a mean temperature below 24°C are considered high-risk (Anis, 2020).
- •19°C: Conducive to the spread, lasting about 60 days (Bu et al, 2020).
- •5°C-15°C: 60% of COVID-19 cases occurred within this range, with a peak at 11.54°C (Huang et al, 2020).
- •1°C increase in temperature leads to a 3.08% reduction in cases (Wu et al., 2020).

B. Humidity Range

- •Above 6 g/kg: Virus transmission is hindered (Kassem, 2020).
- •1% increase in humidity results in a 0.85% reduction in cases (Wu et al., 2020).
- •50%-80% humidity: Suitable for virus survival, with about 75% humidity being ideal for survival (Bu et al, 2020).
- •High humidity with low temperature: Increases COVID-19 transmission (Lin et al, 2020).
- •High humidity with high temperature: Reduces COVID-19 transmission (Lin et al, 2020).
- •Lower humidity with reduced rainfall increases the opportunity for transmission (Bu et al., 2020).

Classifying and ranking the temperature and humidity data

In this section, the researcher extracted and categorized data from previous studies (see Table 10). Microsoft Excel was then used to develop equations that classify the training data for the deep learning framework, based on insights from these studies and expert knowledge.

TABLE 10
EXTRACTED AND CATEGORIZED THE VALUES FOR TEMPERATURE AND HUMIDITY FACTORS

Temperature Range (°C)	Humidity Range (%)	Effect on COVID-19 Spread	Category
5°C - 10°C	50% - 80%	Optimal for transmission, with peak around 10°C	Very High
11°C - 15°C	50% - 80%	High transmission potential, peak around 11.54°C	High
13°C - 19°C	50% - 80%	Conducive for virus survival and transmission	High
13°C - 24°C	<50%	Suitable for virus activity and spread, especially in lower humidity	High
>11°C	>6 g/kg (humidity)	Virus transmission hindered	Low
19°C - 24°C	50% - 75%	Virus survival conducive but reduces with higher temperature	Low
>24°C	Any	Virus spread significantly hindered	Very Low
Any temperature	<50%	Increased transmission in dry conditions	High

- •Very High: Optimal conditions for virus transmission, leading to a rapid spread.
- •High: Suitable conditions for virus survival and spread, though not as severe as peak conditions.
- •Low: The transmission is reduced but still possible.
- •Very Low: Virus transmission is hindered, and survival rates decrease.

TABLE 11
STANDARDIZATION OF THE WEATHER INDICATOR FACTORS

Scale	Weather Data		
	Maximum Temperature (Degree) Range	Minimum Temperature (Degree) Range	Absolute Humidity Range
1	>30		
2	0 To 5 Or >23 <=30	-20 To 0 Or > 16 <=30	0 to 30%
3	>15 <=23	> 7 <= 16	30% to 50%

4	> 5 <= 15	> 0 <= 7	50% to 100%

Standardization table (1-11) shows Vulnerability Scaling (1 = very low, 2 = low, 3 = high, and 4 = very high).

2) ***Data Pre-processing:*** Data pre-processing involves cleaning and transforming raw data into a usable format. The following steps were performed on the dataset:
- Handling Missing Data: Missing values were imputed using mean imputation for continuous variables.
- Normalization: Continuous features, such as temperature and humidity, were normalized to a [0,1] range using Min-Max Scaling to ensure that no single feature dominates the learning process.
- Splitting the Data: The dataset was split into 70% training, 30% Testing subsets to evaluate the model's performance.

A. Deep Learning Architecture

The deep learning framework consists of a feedforward neural network (FNN), designed as follows see fig. 17:

- Input Layer: The input layer accepts a vector of features, including weather data (Maximum temperature, Minimum Temperature and absolute humidity).
- Hidden Layers: The model contains three hidden layers with ReLU activation functions to introduce non-linearity and allow the model to learn complex patterns.

- Layer 1: 128 neurons, ReLU activation
- Layer 2: 64 neurons, ReLU activation
- Layer 3: 32 neurons, ReLU activation
- Output Layer: The output layer consists of a single neuron

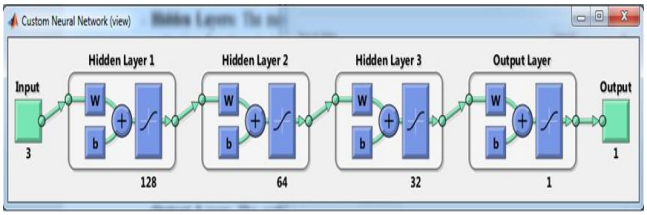

Fig. 17, feedforward neural networks, By Researcher

B. Activation Functions

ReLU (Rectified Linear Unit): ReLU was used for the hidden layers because it helps mitigate the vanishing gradient problem and allows the model to learn faster and more effectively which allow outputs only positive numbers.

C. Optimizer and Loss Function

Optimizer:Levenberg-Marquardt(‘trainlm) optimizer was chosen for its efficiency for small and medium networks in MATLAB and its ability to adjust learning rates dynamically.

Loss Function: Mean Squared Error (MSE) is used as the loss function is regression tasks to measure the average squared differences between the predicted and actual values. It helps in minimizing the error by adjusting the model parameters during training.

D. Model Training

The model was trained using the training data for 74 epochs with a batch size of 32. The model was evaluated on the validation set after each epoch to monitor overfitting and ensure that the model was learning effectively. Best validation Performance is 0.0020753 at epoch 68 suggests the model's predictions are extremely close to the true values, indicating strong performance

E. Evaluation Metrics

Mean Squared Error (MSE): is a widely used metric for evaluating the performance of regression models. It quantifies the difference between the actual (true) values and the predicted values made by the model. MSE is calculated by squaring the differences between each predicted value and its corresponding actual value, then averaging those squared differences.

F. Model Validation

The model was validated using the testing dataset (30% of the data). The test data was used to evaluate the model's generalizability and performance on unseen data.

G. Training and Validation Loss

The training and validation loss curves were plotted to evaluate whether the model was overfitting or under fitting. The training loss decreased steadily, while the validation loss also decreased but showed some fluctuations. This indicates that the model

was learning effectively but still had some room for improvement in generalization.

H. Performance on Test Data

The model achieved the following performance metrics on the test dataset: Accuracy: 93% - MSE = 0.0020753

I. Decision Tree

To benchmark the performance of the neural network, a decision tree classifier was also trained on the same dataset. Decision trees are simple, rule-based models that split the input data based on feature thresholds to make predictions.

J. Evaluation Metrics for decision tree

Accuracy = 87% - Precision: 85% - Recall: 89% - F1-Score: .87

K. Comparison between the performance of Neural Network and Decision Tree

The decision tree classifier achieved an accuracy of 87%, which is slightly lower than the neural network classifier's 93%. While decision trees are easier to interpret and require less computational effort, they may not capture complex, non-linear relationships in the data as effectively as neural networks. However, their simplicity makes them useful as baseline models and for insight into feature importance. Overall, the neural network provided better classification performance in this case, but the decision tree offers a valuable point of comparison.

TABLE 12
PERFORMANCE COMPARISON BETWEEN NEURAL NETWORK AND DECISION TREE

Model	Accuracy (%)	Notes
Neural Network	93	Good generalization
Decision Tree	87	Simpler but slightly less accurate

X. CREATING SPATIAL DISTRIBUTION MAP FOR COVID-19 BASED ON DEEP LEARNING

Using deep learning predictions, a vulnerability map for the study regions was generated through GIS techniques, based on the results of applying deep learning methods. The map categorizes areas into high, moderate, and low vulnerability to epidemic spread, with darker regions indicating higher vulnerability.

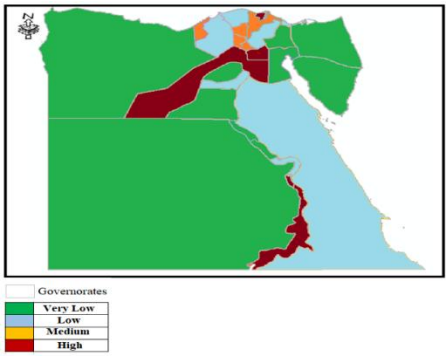

Fig. 18, Scoring for deep learning scenario

XI. SPATIAL DISTRIBUTION MAP FOR COVID-19 BASED ON GIS AND DEEP LEARNING.

Based on the previous approach using GIS and deep learning models, a COVID-19 vulnerability map for Egypt was developed (see Fig 19). The map classifies Egyptian governorates into four COVID-19 vulnerability categories: very low, low, medium, and high, with equal weights assigned to all factors. The analysis was conducted using data from Table 7, Table 8, and Table 9, as well as Fig.16 and Fig.18, applying the overlay intersection tools in GIS ArcMap.

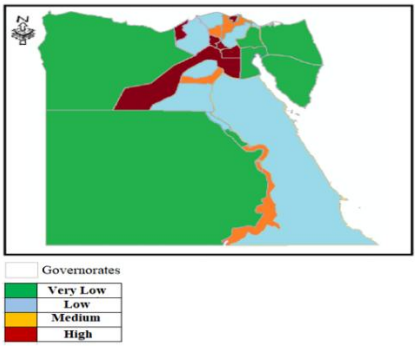

Fig. 19, Scoring for GIS and deep learning scenario.

XII. CONCLUSIONS

The spatial distribution map of COVID-19 risks for Egypt's governorates, developed using GIS and deep learning, was consistent with the official data from the Egyptian Ministry of Health and Population (MOHP). Initially, the GIS-based model predicted that governorates like Cairo and Alexandria were the most vulnerable to COVID-19 spread. However, after integrating GIS model with

deep learning model, the vulnerability map changed, revealing that other governorates, such as Luxor and Qena, had higher predicted risks due to seasonal weather variations in addition to demographic, residential and environmental factors.

The integration of weather data showed that seasonal weather changes played a significant role in the spatial distribution of the epidemic. For instance, warmer months saw a decrease in transmission in coastal regions with higher humidity, while colder months contributed to higher risks in southern governorates like Aswan. The demographic, residential, environmental, and geographical characteristics of these regions further influenced the variation in COVID-19 vulnerability.

Using GIS techniques such as overlay and intersect, the study integrated multiple data layers, including population density, healthcare access, and air quality, to develop a comprehensive vulnerability map. This methodology allowed for a more accurate prediction of high-risk areas, which could be used to inform targeted health interventions.

The results were validated by comparing the prediction scenario with official data from the MOHP. The proposed framework showed a 92% accuracy rate and an 88% precision when compared to reported COVID-19 cases, confirming its reliability as a tool for identifying vulnerable regions.

This research highlights the potential of the proposed framework to support health sector decision-making by identifying areas most at risk for COVID-19 outbreaks. The framework can assist public health authorities in allocating resources more effectively, prioritizing interventions, and optimizing pandemic containment strategies. Future work will focus on incorporating real-time data and improving the model's predictive capabilities.

REFERENCES

[1] WHO 2020, "Coronavirus disease 2019 (COVID-19): situation report", World Health Organization, pp. 72,82

[2] Eales, T., et al. (2020). "Social distancing, population density, and the spread of COVID-19 in urban areas", Health & Place.

[3] Verity R., Okell L.C., Dorigatti I., Winskill P., Whittaker C., Imai N., et al., (2020), "Estimates of the severity of coronavirus disease 2019: a model-based analysis", Lancet Infect. Dis. 20.

[4] Arti, M., Bhatnagar, K., (2020), "Modeling and predictions for COVID 19 spread in India". no. April.

[5] Pandey, G., Chaudhary, P., Gupta, R., Pal, S. ,(2020). "SEIR and Regression Model based COVID-19 outbreak predictions in India". arXiv preprint arXiv:2004.00958.

[6] Dehghan S., Z., Shahnazi, R., (2020). "Spatial distribution dynamics and prediction of COVID-19 in Asian countries: spatial Markov chain approach". Regional Sci. Policy Practice 12, 1005–1025.

[7] Shereen, M.A., Khan, S., Kazmi, A., Bashir, N., Siddique, R., (2020). "COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses". J. Adv. Res. 24, 91–98.

[8] Wang, J., Tang, K., Feng, K., Lv, W., (2020). "High temperature and high humidity reduce the transmission of COVID-19". Available at SSRN 3551767.

[9] Pérez G., Forcadell-Díez L., Reyes A., Perez C., Bartoll X., Borrell C.,(2024)," Housing conditions and COVID-19 in Barcelona: do they change by gender?" BMC Public Health ,(BioMed Central) , - Vol. 24, Iss: 1.

[10] Coşkun H., H., Yildirim, N., Gunduz, S., (2021). "The spread of COVID-19 virus through population density and wind in Turkey cities". Sci. Total Environ. 751, 141663.

[11] Saiman L.,Edward E. Walsh,Angela R. Branche, Barrett A., Alba L, Gollerkeri S., Julia A. Schillinger, Phillips M., Finelli L.,(2024)."Impact of Age and Comorbid Conditions on Incidence Rates of COVID-19-Associated Hospitalizations, 2020–2021", Influenza and Other Respiratory Viruses (Wiley) - Vol. 18, Iss: 11.

[12] Bayram H,Konyalilar N,Elci M. A., Rajabi H., Aksoy G. T., Mortazavi D., Kayalar Ö., Dikensoy Ö, Taborda-Barata L., Viegi G.,(2024)."Issue 4-Impact of air pollution on COVID-19 mortality and morbidity: An epidemiological and mechanistic review", Pulmonology

[13] Paulo R. Martins-Filho ,(2021)." Relationship between population density and COVID-19 incidence and mortality estimates: A county-level analysis", PubMed Central, PMC8253654, PMID: 34245973.

[14] Kamis C., et al.,(2021)," Overcrowding and COVID-19 mortality across U.S. counties: Are disparities growing over time?", SSM - Population Health, Volume 15.

[15] Costa, Solange, Ferreira, Joana, Silveira, Carlos, Costa, Carla, Lopes, Diogo, Relvas, Hélder, Borrego, Carlos, Roebeling, Peter, Miranda, Ana Isabel, Paulo Teixeira, João, (2014). "Integrating health on air quality assessment-review report on health risks of two major European outdoor air pollutants: PM and NO2." J. Toxicol. Environ. Health Part B 17 (6), 307–340. C.

[16] Liu M., Li Z., Liu M., Zhu Y., Liu Y., Kuetche M.,Wang J., Wang X., Liu X., Li X., Wang W., Guo X.,(2022). "Association between temperature and COVID-19 transmission in 153 countries", Springer, Volume 29, pages 16017– 16027.

[17] Kassem, A. Z., (2020)," Does Temperature Affect COVID-19 Transmission?", PubMed Central, National Library of Medicine, PMCID: PMC7793668, PMID: 33425828

[18] Saaty, T.L., (1977). "A scaling method for priorities in hierarchical structures". J. Math.Psychol. 15, 234–281.Sarwar, Adnan, Imran, Muhammad,

[19] Singh, Rohit, Avikal, Shwetank, (2020). "COVID-19: A decision-making approach for prioritization of preventive activities". Int. J. Healthcare Manage. 13 (3), 257– 262.

[20] Saaty T.L.,(1980). "The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation", McGraw-Hill International Book Co., New York

[21] Barczewski J. ,(1999)." GIS and Multi-Criteria Decision Analysis", Wiley, New York, p. 392.