Surface Web Merits for SARS-CoV-2 Pandemic in Iraq

DOI: https://doi.org/ 10.32007/jfacmedbagdad.6241795

Ahmed Al-Imam* **MRes**
Marek A. Motyka** **PhD**
Hend J. Al-Doori*** **CABMS**

JFac Med Baghdad 2020; Vol.62, No. 4
Received: Oct. 2020
Accepted: Dec. 2020
Published: Feb. 2021

Abstract

Background: Data on SARS-CoV-2 from developing countries is not entirely accurate, demanding incorporating digital epidemiology data on the pandemic.

Objectives: To reconcile non-Bayesian models and artificial intelligence connected with digital and classical (non-digital) epidemiological data on SARS-CoV-2 pandemic in Iraq.

Methods: Our study design is longitudinal, for the period from 24 February 2020 to 25 September 2020. We retrieved data from the Iraqi Ministry of Health on the daily cases, recoveries, and deaths from SARS-CoV-2, and incorporated collateral data from Google Trends using five search terms, "Coronavirus", "كورونا", "COVID-19", "كوفيد-19", and "لقاح كورونا". The search terms "كورونا", "كوفيد-19", and "لقاح كورونا" represent the Arabic translations for "Coronavirus", "COVID-19", and "COVID-19 Vaccine". We implemented multivariate tests and machine learning to scrutinize the spatio-temporal trends of the pandemic in Iraq and interpret the causality influencing Iraqis to seek digital knowledge, via the web, on SARS-CoV-2.

Results: Baghdad and Sulaymaniyah represented statistical outliers in connection with daily cases and recoveries, and daily deaths, respectively. Multivariate tests and neural networks detected a predictor effect of deaths, recoveries, and daily cases on web searches concerning two search terms, "كورونا" and "Coronavirus" (Pillai's Trace value=1, F=1106915.624, Hypothesis df=3, Error df=12, p-value<0.001, Partial Eta Squared=1). Using hierarchical clustering, we identified distinctive aggregates involving the Iraqi capital, Kurdistan region, and the south of Iraq. Three search terms were most prevalent among Iraqi web users, including "كورونا", "كوفيد-19", and "Coronavirus". Significant bivariate correlations were all positive except for those involving the search term "لقاح كورونا". Al-Muthanna governorate residents were least interested in data on SARS-CoV-2 vaccines.

Conclusion: Our analyses were triumphant in syncretizing non-Bayesian and machine learning models, using two forms of epidemiology data on the pandemic in Iraq. We opine that the current study is exquisite and precious for decision-makers at the Iraqi Ministry of Health.

Keywords: Artificial intelligence; coronaviridae; COVID-19; digital epidemiology; epidemiology; internet; machine learning; novel coronavirus; SARS-CoV-2; Surface web.

Introduction:

The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), which is responsible for the Coronavirus disease 2019 (COVID-19), is a virulent pathogenic zoonotic viral infection [1]. By the end of 2019, COVID-19 started in the city of Wuhan in China, in which the Chinese Centre for Disease Control and Prevention reported numerous ambiguous cases of pneumonia of an unknown aetiology among the Wuhan residents [1]. Microbiologists identified a novel coronavirus based on clinical samples, and they sequenced its genome [2]. SARS-CoV-2 infections occur due to close contact with an infected person who, by coughing, sneezing, or merely breathing, excretes virus-laden aerosols infective to others [1, 2]. COVID-19 manifests as fever and other constitutional clinical features, including chills, fatigue, and muscle aches [3]. Later, the patient develops dry cough and dyspnea, while rhinitis and sore throat are rare, and roughly up to 25% of patients may develop diarrhoea later [1, 3]. During the early phase of COVID-19, some individuals may develop a loss of smell and taste [1, 3]. In Wuhan, almost one-third of patients required intensive care hospitalization, and relatively a high mortality (10%), while higher mortality rates (50%) existed in patients over the age of 50 years, in whom death usually occurs during the third week from the onset of symptoms [1, 3]. SARS-CoV-2 manifestations were more severe in patients with pre-existing medical and surgical conditions, including cardiovascular and immunological disturbances [1-3]. Since the Wuhan outbreak's initial reports, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has been spreading worldwide at an alarmingly exponential rate [1, 2]. However, on 18 November 2020 and based on data

**Department of Anatomy, College of Medicine, University of Baghdad. email:ahmed.lutfi@uobghdad.edu.iq*
*** Institute of Sociology, University of Rzeszów, Poland.email:mmotyka@ur.edu.pl marek.a.motyka@tlen.pl*
**** Al-Betool Teaching Hospital, Diyala Health Directorate, Iraqi Ministry of Health. email: hendaldoori0571@gmail.com*

from serology study using specific antibodies for the virus in lung cancer patients who were exposed to the virus, the South China Morning Post claimed the diseased might have originated initially in Italy as early as September 2019 [3]. As of 20 November 2020, the number of confirmed infections exceeded 57,236,335 worldwide, and 529,226 cases in Iraq [4]. Complications related to the illness claimed the lives of over 1,365,634 globally, and 11,834 individuals in Iraq [4]. SARS-CoV-2 allocated with to nations from the developed and the developing world as well, including the United States, India, Brazil, Russia, Columbia, Peru, Spain, Mexico, Argentina, South Africa, France, Chile, Iran, UK, Bangladesh, Iraq, Saudi Arabia, Turkey, Italy, and Pakistan [4]. As a top priority for the global health agenda, researchers are developing effective vaccines in several nations globally, including China, Russia, the United Arab Emirates, the United Kingdom, and the United States [5-7]. There are currently more than 180 vaccines in development, some of which have moved into phase III clinical trials, and holistic data on the vaccines' effectiveness and safety will be available within few months [6]. Two vaccines lead the race, Pfizer's and Moderna's, with 95% and 94.5% effectiveness, respectively [7].

The present study examines the Iraqi situation on SARS-CoV-2 by implementing non-Bayesian models and machine learning analytics to contrast digital and classical epidemiological data for a summative period of seven months starting from the first documented case of COVID-19 infection in the Iraqi population. Our primary objective is to examine the pandemic spatio-temporal patterns in Iraq while explaining the factors influencing web searches connected with SARS-COV-2, using polynomial regression, multivariate tests, and machine learning models. Our null hypothesis is that there will be no disparity between digital and non-digital epidemiological data. By incorporating predictive analytics to anticipate upcoming waves and seasonal variabilities of the pandemic in Iraq, we opine that the current study will be of prime importance for decision-makers at the Iraqi Ministry of Health by, and provide novel insights on the importance of digital knowledge existing on the surface web to predict future trends in the SARS-CoV-2 pandemic, using techniques of artificial intelligence [8-11].

Materials and Methods:

Ethics and Study Design: The study was conducted following the standard protocol of ethics and scientific committee of the College of Medicine at the University of Baghdad, the declaration of Helsinki by World Medical Association, the EU protocol on protection of animals used for scientific purposes (EU Directive 210/63/EU), and the ethical principles of Framingham consensus of 1997. We based our study on snapshots that we took for the surface web, via Google search engine while using Google Chrome web browser [version 85.0.4183.121 (Official Build) (64-bit)]. We took snapshots for Google Trends and compared them with real epidemiological data from the Iraqi Ministry of Health. Our study represents a longitudinal data analytic from 24 February 2020 to 25 September 2020. The date of 24 February 2020 represents the time of discovering the first case of COVID-19 in Iraq, also known as patient zero [12].

Digital and Classical Epidemiological Data Google Trends and the Iraqi Ministry of Health

To map the geographic and temporal distribution of the interest, i.e., the spatial and temporal variabilities of the search volume, of the surface web users who searched for data in connection with the SARS-CoV-2 pandemic, we used Google Trends to retrieve a retrospective longitudinal data for the period we specified earlier. We took the snapshot for five search terms, including "Coronavirus", "كورونا", "COVID-19", "كوفيد-19", and "لقاح كورونا" [13]. The three search terms "كورونا", "كوفيد-19", and "لقاح كورونا" represent the Arabic translations for "Coronavirus", "COVID-19", and "COVID-19 Vaccine". We analyzed the search volume (number of hits) of those search terms in conjunction with classical epidemiological data from the Iraqi Ministry of Health that we retrieved from Worldometer website using a web scraper tool and a dedicated script written in Python high-level programming language [4, 14].

Statistical Analysis, and Level-of-Evidence:

We used the IBM Statistical Package for the Social Sciences (IBM-SPSS version 24) and Microsoft Office Excel 2016 with integrated Data Analysis ToolPak add-in. For hypothesis testing, we adopted a p-value of 0.05 as the cut-off margin for statistical significance. We are using a hybrid of analytics for two main reasons; to detect potential statistically significant findings of small effect size, and to provide a collateral confirmatory evidence for those tests. The different data analytics are not alternatives, but complementary to each other, serving a summative "grand" analysis of the pandemic in our country. On the other side, predictive analytics will serve the sole purpose of anticipating upcoming (future) digital trends in connection with the pandemic in Iraq, with an aim to orient Iraqi decision-making policy concerning the pandemic. For instance, how to mobilize and allocate the Ministry of Health resources from one Iraqi governorate to the other, based on predictive analytics. In parallel with the bivariate and multivariate statistical models, we shall run supervised machine learning, in the form of multilayer perceptron neural networks, using scaled conjugate gradient optimization algorithm, and a default SPSS allocation of the training set and testing set at 70% and 30%, respectively [15]. The neural networks will yield synaptic weights and independent variables importance analysis equivalent to the effect size in non-Bayesian statistical analysis. Unsupervised machine learning, using cluster analysis, will complement our summative statistical and neural network analytics. Principally, we will use hierarchical cluster analysis to identify potential aggregates clusters within our sample [16]. Finally, we evaluated the level-of-evidence according to the categorization scheme of the Oxford Centre for Evidence-based Medicine [17].

Results:

Temporal Mapping

Descriptive Statistics and Tests of Normality

Three search terms had the highest interest among surface web users from Iraq, including “Coronavirus” (mean=13.44, standard error=0.597, skewness=5.277, kurtosis=47.290), “كورونا” (8.54, 0.418, 9.848, 122.711), and “كوفيد-19” (2.30, 0.432, 3.467, 10.334). These three search terms also had the highest temporal variations (dispersion over time) (Figure 1). Further, statistical outliers, for "Coronavirus" and "كورونا", existed on the 24th and the 25th of February 2020, i.e. around the time when patient zero appeared in Iraq, while statistical outliers for the three search terms also appeared later on April 2020 between 1 April 2020 and 19 April 2020. Two search terms had the least interest among surface web users from Iraq, including “COVID-19” (mean=0.95, standard error=0.015, skewness=-4.279, kurtosis=16.468), and “لقاح كورونا” (0.79, 0.028, -1.403, -0.033). These two search terms also had the least dispersion over time and had no statistical outliers. Accordingly, Iraq surface web users were most interested in three search terms, two of which were in the Arabic language. Hence, we shall select "Coronavirus", "كورونا", and "كوفيد-19" search terms as dependent variables in response to time as the independent variable, for consequent data models. According to Shapiro-Wilk test of normality, none of the five search terms possessed a normal distribution, including “Coronavirus” (test statistic=0.613, df=210, p-value<0.001), “كورونا” (0.384, 210, p<0.001), “COVID-19” (0.217, 210, p<0.001), “كوفيد-19” (0.344, 210, 0.001), and “لقاح كورونا” (0.505, 210, p<0.001). Therefore, we shall conduct nonparametric correlations using Kendall's Tau bivariate correlations.

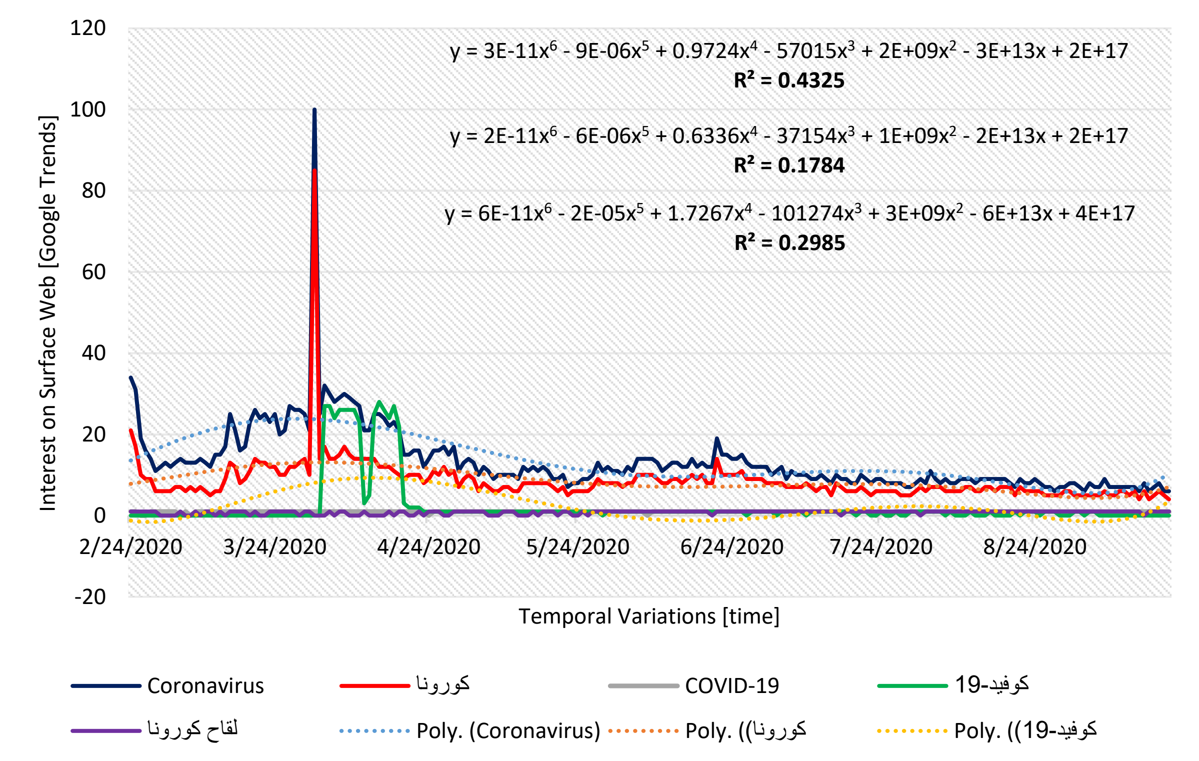

Figure 1: Temporal Mapping of Search Terms on SARS-CoV-2 Pandemic in Iraq.

* The equations relate to the most popular three search terms (Coronavirus, كورونا, and كوفيد-19). ** x=time, y=predicted number of searches on the surface web.

Kendall's Tau Bivariate Correlations

Significant bivariate correlations existed for "كورونا" versus "Coronavirus" (Kendall's tau_b=0.823, p-value<0.001), "كورونا" versus "كوفيد-19" (0.278, p<0.001), "كورونا" versus "لقاح كورونا" (-0.170, p=0.004), "كوفيد-19" versus "Coronavirus" (0.180, p=0.001), "كوفيد-19" versus "COVID-19" (0.178, p=0.007), and "كوفيد-19" versus "لقاح كورونا" (-0.134. p=0.044), and "لقاح كورونا" versus "Coronavirus" (-0.249, p<0.001). To summarize, all significant correlations were positive except for those including the search term "لقاح كورونا", and all significant correlations had a weak effect size excluding "كورونا" versus "Coronavirus" which had a strong effect size. For seven months following the discovery of patient zero, Iraqis were most interested in searching the Internet for data on SARS-CoV-2 pandemic using two strongly correlated search terms, "Coronavirus" and "كورونا", while being least interested, in searching the web for data on potential vaccines that may end the pandemic.

Multivariate Tests and Predictive Analytics

Here, we deployed a multivariate analysis of variance (MANOVA) while considering time as an independent variable (predictor) and the search terms as dependent variables (outcomes). According to the multivariate model (Pillai's Trace value=0.972, F=1363.313, Hypothesis df=5, Error df=199, p-value<0.001, Partial Eta Squared=0.972), time was a strong significant predictor influencing Iraqis to search the surface web for data relevant to SARS-CoV-2, and they were most motivated to seek web-based knowledge during February and April 2020. Therefore, using polynomial regression, we created three predictive models, one for each of the three most popular search terms (x, dependent variable) and as a function of time (y, independent variable) (Table 1). The three models were successful, each of which had a medium (moderate) effect size, including "Coronavirus" ($y = 3E\text{-}11x^6 - 9E\text{-}06x^5 + 0.9724x^4 - 57015x^3 + 2E{+}09x^2 - 3E{+}13x + 2E{+}17$, $R^2=0.4325$, Correlation Coefficient=0.6577), "كوفيد-19" ($y = 6E\text{-}11x^6 - 2E\text{-}05x^5 + 1.7267x^4 - 101274x^3 + 3E{+}09x^2 - 6E{+}13x + 4E{+}17$, $R^2=0.2985$, Correlation Coefficient=0.5464), and "كورونا" ($y = 2E\text{-}11x^6 - 6E\text{-}06x^5 + 0.6336x^4 - 37154x^3 + 1E{+}09x^2 - 2E{+}13x + 2E{+}17$, $R^2=0.1784$, Correlation Coefficient=0.4224). Each of these models can inform decision-making of Iraqi health officials by predicting future changes in digital epidemiology, based on a specific search term, with time.

Table 1. Predictive Models of the Most Popular Search Terms on SARS-CoV-2 Pandemic in Iraq.

Search Term *	Equation of the Predictive Model **	Order of Polynomial Function	R^2 Score	Correlation Coefficient	Effect Size
Coronavirus	$y = 3E\text{-}11x^6 - 9E\text{-}06x^5 + 0.9724x^4 - 57015x^3 + 2E{+}09x^2 - 3E{+}13x + 2E{+}17$	6th	0.4325	0.6577	Moderate
كوفيد-19	$y = 6E\text{-}11x^6 - 2E\text{-}05x^5 + 1.7267x^4 - 101274x^3 + 3E{+}09x^2 - 6E{+}13x + 4E{+}17$	6th	0.2985	0.5464	Moderate
كورونا	$y = 2E\text{-}11x^6 - 6E\text{-}06x^5 + 0.6336x^4 - 37154x^3 + 1E{+}09x^2 - 2E{+}13x + 2E{+}17$	6th	0.1784	0.4224	Moderate

* Date: 24.02.2020 to 25.09.2020. ** x=time, y=predicted number of searches on the surface web.

Spatial (Geographic) Mapping

Descriptive Statistics and Tests of Normality

We took a snapshot, via Google Trends, on 25 September 2020 for the same five search terms to assess the geographic mapping of each term for all the Iraqi governorates, and to compare them with classical epidemiological data on the daily cases, recoveries, and daily deaths from SARS-CoV-2 as confirmed by the Iraqi Ministry of Health. The search terms "Coronavirus", "كورونا", and "كوفيد-19" had statistical outliers only in connection with three Iraqi governorates, including Duhok, Erbil, and Sulaymaniyah. Iraqis from the region of Kurdistan used two search terms the least, "كورونا" and "كوفيد-19", while using the English alternative, "Coronavirus", to browse the web for digital knowledge on the pandemic (Figure 2). Further, Iraqis from Al-Muthanna Governorate showed no interest in browsing the web for data on potential vaccines, while their fellow citizens from Dhi-Qar governorate showed the highest interest to seek digital-based knowledge on potential vaccines for SARS-CoV-2 (Figure 3). On the other hand, data from the Ministry of Health displayed statistical outliers connected with daily recoveries and daily new cases in Baghdad, and daily deaths in Sulaymaniyah. Again, all of the parameters, including the search terms, daily cases, recoveries, and daily deaths, were not normally distributed.

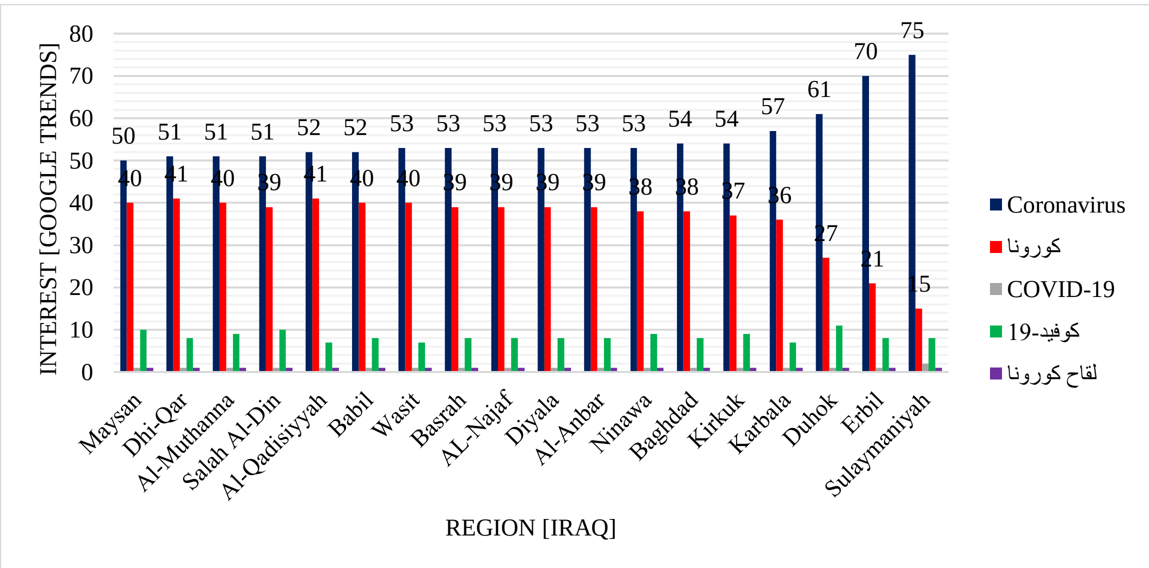

Figure 2: Spatial Mapping of Search Terms on SARS-CoV-2 Pandemic in Iraq.

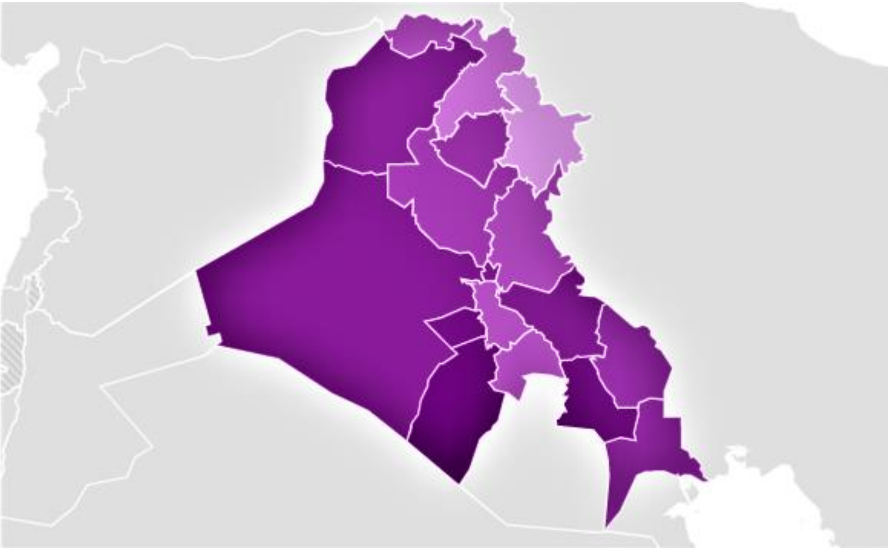

Figure 3. Geographic (Spatial) Mapping of the Search Term "كورونا لقاح" in Iraq.

Table 2. Bivariate Correlations Matrix.

		Coronavirus	كورونا	COVID-19	كوفيد-19	كورونا لقاح	Recoveries	New Cases	Deaths
Coronavirus	Correlation Coefficient	1.000	-.813**	.358	-.165	.	.063	.189	.290
	Sig. (2-tailed)	.	.000	.094	.402	.	.727	.294	.123
	N	18	18	18	18	18	18	18	18
كورونا	Correlation Coefficient	-.813**	1.000	-.355	-.148	.	.090	-.063	-.303
	Sig. (2-tailed)	.000	.	.096	.453	.	.615	.728	.107
	N	18	18	18	18	18	18	18	18
COVID-19	Correlation Coefficient	.358	-.355	1.000	-.069	.	-.216	.059	.354
	Sig. (2-tailed)	.094	.096	.	.756	.	.289	.772	.096
	N	18	18	18	18	18	18	18	18
كوفيد-19	Correlation Coefficient	-.165	-.148	-.069	1.000	.	-.216	-.139	.041
	Sig. (2-tailed)	.402	.453	.756	.	.	.252	.461	.835
	N	18	18	18	18	18	18	18	18
كورونا لقاح	Correlation Coefficient	.	.	.	.	.	.	.	.
	Sig. (2-tailed)	.	.	.	.	.	.	.	.
	N	18	18	18	18	18	18	18	18
Recoveries	Correlation Coefficient	.063	.090	-.216	-.216	.	1.000	.386*	.263
	Sig. (2-tailed)	.727	.615	.289	.252	.	.	.025	.143
	N	18	18	18	18	18	18	18	18
New Cases	Correlation Coefficient	.189	-.063	.059	-.139	.	.386*	1.000	.333
	Sig. (2-tailed)	.294	.728	.772	.461	.	.025	.	.064
	N	18	18	18	18	18	18	18	18
Deaths	Correlation Coefficient	.290	-.303	.354	.041	.	.263	.333	1.000
	Sig. (2-tailed)	.123	.107	.096	.835	.	.143	.064	.
	N	18	18	18	18	18	18	18	18

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Only one significant bivariate correlation, of positive and moderate effect size, existed within classical epidemiological data, including daily new cases versus recoveries (Kendall's tau_b=0.386, p-value=0.025) (Table 2). Within digital epidemiological data, using search terms, also one significant correlation of a strong effect size existed between "Coronavirus" and "كورونا" (-0.813, p<0.001). No significant correlations linked digital and classical epidemiological data.

Multivariate Tests

Here, we deployed another multivariate analysis of variance (MANOVA) while considering SARS-CoV-2 daily cases, recoveries, and daily deaths as independent variables (predictors) and the three search terms, "Coronavirus", "كورونا", and "كوفيد-19" as dependent variables (outcomes). According to the multivariate tests (Pillai's Trace value=1, F=1106915.624, Hypothesis df=3, Error df=12, p-value<0.001, Partial Eta Squared=1), the predictors within the overall corrected model had a significant effect over the search term "كورونا" (df=3, F=6.366, p-value=0.006, Partial Eta Squared=0.577). Nonetheless, only one of the predictor, deaths from COVID-19, significantly and strongly influenced Iraqis to browse the Internet seeking digital knowledge on SARS-CoV-2 using two search terms, "كورونا" (df=1, F=14.740, p-value=0.002, Partial Eta Squared=0.513) and "Coronavirus" (df=1, F= 13.315, p=0.003, 0.487) (Table 3).

Table 3. Multivariate Tests: Tests of Between-Subjects Effects.

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Coronavirus	419.562[a]	3	139.854	5.432	.011	.538
	كورونا	527.919[b]	3	175.973	6.366	.006	.577
	كوفيد-19	2.271[c]	3	.757	.589	.633	.112
	Coronavirus	14638.102	1	14638.102	568.567	.000	.976
	كورونا	9094.906	1	9094.906	328.993	.000	.959
	كوفيد-19	418.957	1	418.957	325.730	.000	.959
Recoveries	Coronavirus	22.448	1	22.448	.872	.366	.059
	كورونا	42.964	1	42.964	1.554	.233	.100
	كوفيد-19	1.906	1	1.906	1.482	.244	.096
New Cases	Coronavirus	14.297	1	14.297	.555	.468	.038
	كورونا	27.681	1	27.681	1.001	.334	.067
	كوفيد-19	1.352	1	1.352	1.051	.323	.070
Deaths	Coronavirus	342.805	1	342.805	13.315	.003	.487
	كورونا	407.490	1	407.490	14.740	.002	.513
	كوفيد-19	.001	1	.001	.000	.984	.000
Error	Coronavirus	360.438	14	25.746			
	كورونا	387.025	14	27.645			
	كوفيد-19	18.007	14	1.286			
Total	Coronavirus	55892.000	18				
	كورونا	24315.000	18				
	كوفيد-19	1287.000	18				
Corrected Total	Coronavirus	780.000	17				
	كورونا	914.944	17				
	كوفيد-19	20.278	17				

a. R Squared = .538 (Adjusted R Squared = .439)
b. R Squared = .577 (Adjusted R Squared = .486)
c. R Squared = .112 (Adjusted R Squared = -.078)

Supervised and Unsupervised Machine Learning

To reconcile machine learning and non-Bayesian models, we deployed a neural network of the same architecture, concerning predictors and outcomes, of the previous multivariate model (Table 4, Figure 4). The network yielded comparable results, assigning the independent variable importance to daily new cases (importance=0.553, normalized importance=100%), recoveries (0.318, 57.50%), and deaths (0.13, 23.50%). Those predictors influenced Iraqis' behaviour over the web to seek knowledge connected with the pandemic. To complement machine learning, we used hierarchical cluster analysis to identify potential aggregates of the Iraqi governorates concerning digital and classical epidemiological data. When setting the maximum number of clusters to two, Baghdad occupied one of the clusters while the rest of the governorates occupied the second (Figure 5). When setting the maximum number of clusters to three, Baghdad allocated to one of the clusters, and three neighbouring governorates of the south, Dhi-Qar, Wasit, and Basrah, allocated to the second, while the rest of Iraq allocated to the last cluster. There were similar results when setting the maximum number of clusters to four. However, when setting the maximum number of clusters to five, the first cluster has Baghdad, the second cluster has Dhi-Qar only, the third cluster had Al-Qadisiyyah, Duhok, and Erbil, the fourth cluster had Wasit and Basrah, while the fifth cluster had the rest of Iraq. It appears the SARS-CoV-2 affects Iraq in a unique spatial pattern with a varying predilection for the capital city compared to the south of Iraq and the region of Kurdistan. Perhaps, reflecting an abundance of unknown factors, including ethnicities, climatic differences, and population demographics, interacting with each other to manifest as this pattern of distinct clustering, which necessitates future research.

Table 4. Neural Network Information.

Input Layer	Covariates	1 Recoveries
		2 New Cases
		3 Deaths
	Number of Units[a]	3
	Rescaling Method for Covariates	Standardized
Hidden Layer(s)	Number of Hidden Layers	1
	Number of Units in Hidden Layer 1[a]	2
	Activation Function	Hyperbolic tangent
Output Layer	Dependent Variables	1 Coronavirus
		2 كورونا
		3 كوفيد-19
	Number of Units	3
	Rescaling Method for Scale Dependents	Standardized
	Activation Function	Identity
	Error Function	Sum of Squares

a. Excluding the bias unit

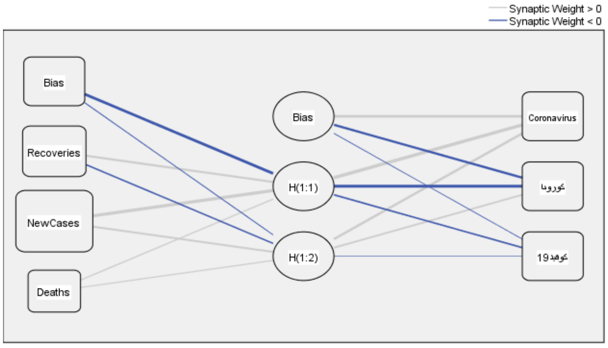

Figure 4. Supervised Machine Learning: Neural Network Analysis.

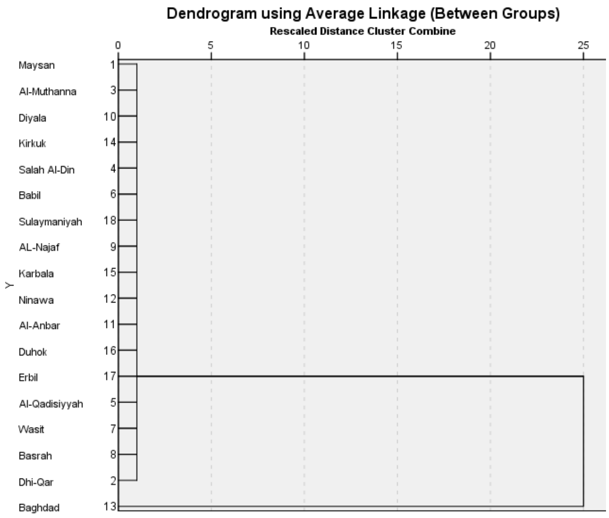

Figure 5. Unsupervised Machine Learning: Dendrogram of Hierarchical Cluster Analysis.

Discussion:

Three search terms were most prevalent among Iraqi users of the surface web, including "Coronavirus", "كورونا", and "كوفيد-19", who browsed the Internet mainly during the beginning of April 2020, and to a less extent around the time of discovery of patient zero in Iraq. Iraqis from the region of Kurdistan used two search terms the least, "كورونا" and "كوفيد-19" while using the English alternative, "Coronavirus", to browse the web for digital knowledge on the pandemic, which is attributable to their native (Kurdish) language restrictions. Contrary to residents of Dhi-Qar, those from Al-Muthanna Governorate were least interested in data on SARS-CoV-2 vaccines. Further, significant bivariate correlations were all positive except for correlations involving the search term "لقاح كورونا". Al-Muthana citizens may have unique cultural and sociodemographic backgrounds that influence their exposure to specific news outlets, media networks, and social media networks. Multivariate tests and polynomial predictive models indicated a significant effect of time on Iraqis to search the web using specific terms on SARS-CoV-2 pandemic. Baghdad and Sulaymaniyah represented statistical outliers in connection with daily cases and recoveries, and daily deaths, respectively. Significant bivariate correlations linking digital and classical epidemiological data were lacking. Nonetheless, multivariate tests and artificial neural networks detected a confirmatory predictor effect of deaths, recoveries, and daily cases on web searches concerning two terms, "كورونا" and "Coronavirus". Hierarchical clustering identified peculiar aggregates involving the Iraqi capital, Kurdistan region, and the south of Iraq, which mandates subsequent scrutiny. It appears the SARS-CoV-2 affects Iraq in a unique spatial pattern with a varying predilection for the capital city compared to the south of Iraq and the region of Kurdistan. Perhaps, reflecting an abundance of explanatory variables, including ethnicities, climatic differences, and population demographics, complexly interacting with each other to manifest as this pattern of distinct clustering, which necessitates future research. There have been numerous rumours and speculations surrounding the pandemic, some of which belongs to the realm of pseudoscience and conspiracy theories. Nevertheless, there was an article published in Nature News in 2015, in which the author debated a "lab-made coronavirus related to SARS can infect human cells" [18, 19]. Regardless of the essence of the novel coronavirus, whether crafted by nature, man, or a super-intelligent extraterrestrial Kardashev civilization connected with the Zoo hypothesis and the Fermi paradox, the "hunt" to find patient zero has been fruitless [8, 20, 21]. Nonetheless, Donna Lu (2020) discussed a study of the first forty-one people who contracted SARS-CoV-2, the first patient was a male who showed symptoms on 1 December 2019, and he had no links to the infamous Huanan Seafood Market [20]. On 24 February 2020, the World Health Organization Regional Office for the Eastern Mediterranean (WHO-EMRO) announced that patient zero in Iraq was an Iranian student who lived in Al-Najaf governorate [22]. Later, BBC News reported the first incident of death in Iraq from SARS-CoV-2 on 4 March 2020, a man who is sixty-three years old from Suleymaniyeh [23]. Our temporal mapping of the surface web detected statistical outliers during February and April 2020, and these coincide with the emergence of the pandemic in Iraq as reported by the WHO-EMRO, as well as the nationwide lockdown that the Iraqi authorities implemented on 22 March 2020 until the mid-April of the same year [22, 24, 25]. As of 5 October 2020, and according to Worldometer website and COVID-19 application (iOS version 0.9.16), the number of daily cases and deaths increased all over the Middle East [4, 26]. The neighboring countries of Iraq also witnessed an exponential growth of SARS-CoV-2 infections, including Iran (total cases=475,674, total deaths=27,192, total recoveries=392,293, mortality rate=6.48%, basic reproduction number=1.07), Kuwait (cases=107,592, deaths=628, recoveries=99,549, MR=0.58%, BRN=0.92), Saudi Arabia (cases=336,766, deaths=4,898, recoveries=322,055, MR=1.45%, BRN=0.88), Jordan (cases=17,464, deaths=110, recoveries=5,292, MR=0.63%, BRN=1.81), Syria (cases=4,411, deaths=207, recoveries=1,168, MR=4.69%, BRN=0.99), and Turkey (cases=326,046, deaths=8,498, recoveries=286,370, MR=2.61%, BRN=0.93) [4, 26]. Unfortunately, Iraq is in the lead of its Arab neighbours in connection with the stats of the pandemic (cases=382,949, deaths=9,464, recoveries=312,158, MR=2.47%, BRN=1.01) [4, 26]. The current study has some limitations, including those of the statistical analyses. For instance, the immoderate type-1 statistical error can manifest as a result of carrying out multiple data analytics. Some tests, including correlation analytics, were more conservative than others as in the Kendall rank correlation. Additionally, the interpretation of causality in our hypotheses and different models that we implemented may accept different interpretations from a philosophical perspective, including arguing the basis of the Bradford Hill criteria on causality relationships when classifying specific variables into independent (predictors) and dependent (outcomes) [27]. Besides, the multivariate tests have inherent limitations of their own as per the renowned British statistician George Box's aphorism, "All models are wrong, but some are useful" [28]. Hence, optimizing statistical models and implementing them for real-time temporal analyses is valuable for future research [29-31]. There are also implicit constraints of the statistical packages, including IBM SPSS and Microsoft Excel, when loading a specific type or a count of variables into a data model, including the multivariate analysis of variance, supervised neural networks, and cluster analysis. Future research can consider the intricate interaction of an abundance of variables, covariables, and cofactors by using advanced statistical packages running on a powerful supercomputer. The coronavirus pandemic has become a significant problem for communities around the world. The challenges posed by the continuous growth of infections and the lack of effective methods to prevent the disease, justify undertaking all activities that can help regain control over the ever-growing threat. Unfortunately, already during a study conducted in spring 2020, it was found that a particular part of the population did not adhere to the safety rules, making it a high-risk group for SARS-CoV-2 infection [1]. Also, conspiracy theories about the coronavirus that disinforms recipients have appeared, especially on the Internet [32]. These contents may significantly contribute to underestimating reliable information from the coronavirus fight's front. The reason may also be the lack of critically evaluating information obtained through electronic media and the transmission of information filtered by government websites or financial institutions to avoid panic [33]. Almost a year after identifying the first infections, scepticism about infection prevention principles, via wearing masks and social distancing, seems to be much greater than a few months ago [34-35]. Bayram and colleagues are calling for a large-scale implementation of innovation policy, which is one of the ways of collective decision making, and provides an opportunity to include a variety of departments in the public policy space that can contribute to the shaping of alternative influences as new technologies and scientific fields emerge, but also new alternative ideas that allow for better monitoring to help predict the development of the pandemic [36]. For months, actions have also been taken to monitor new infections and people under quarantine to stop the spread of new infections [37-40]. Activities that use artificial intelligence to monitor and predict the pandemic's development can be essential to counteract the disease or direct harm reduction to regions, particularly at risk of new outbreaks [36, 41].

Conclusion:
Our analyses were triumphant in syncretizing non-Bayesian and machine learning models, using two forms of epidemiology data on the pandemic in Iraq. We opine that the current study is exquisite and precious for decision-makers at the Iraqi Ministry of Health. The forecasting method proposed by the authors of this article is one of the innovative actions of this kind, which aims to predict and direct preventive actions in the regions of Iraq where the threat is forecasted and, to a broader extent, the welfare of the whole Iraqi society.

Availability of Data
Our data, including the raw dataset, are available upon request from the corresponding author.

Conflict of Interest
The authors declare that they have no conflict of interest and have self-funded this study.

Authors' Contribution

Ahmed Al-Imam collected the data, conducted data analytics, wrote the first draft of the article, and prepared the manuscript for scholarly submission. Marek A. Motyka and Hend J. Al-Doori contributed to developing the study concept, reviewing the first draft and developing the paper for academic presentation.

References
1. Motyka MA, Al-Imam A, Aljarshawi MHA. SARS-CoV-2 pandemic as an anomie. Przestrzeń Społeczna (Social Space). 2/2020 (20): 111-144.
2. Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. Journal of the American Medical Association. 2020; 323(18):1843–1844.
3. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next?. The Lancet. 2020;395(10231):1225-1228.
4. Worldometer - real time world statistics [Internet]. Worldometer. 2020 [cited 29 September 2020]. Available from: https://www.worldometers.info/
5. Amanat F, Krammer F. SARS-CoV-2 Vaccines: Status Report. Immunity. 2020; 52(4): 583-589.
6. Thomas K. New Pfizer Results: Coronavirus Vaccine Is Safe and 95% Effective. Available from https://www.nytimes.com/2020/11/18/health/pfizer-covid-vaccine.html *[cited 20 November 2020].*
7. Plebanski M, Apostolopoulos V. Moderna's COVID vaccine reports 95% efficacy. It means we

might have multiple successful vaccines. Available from https://theconversation.com/modernas-covid-vaccine-reports-95-efficacy-it-means-we-might-have-multiple-successful-vaccines-150266 [cited 20 November 2020].
8. Al-Imam A, Motyka MA, Jędrzejko MZ. Conflicting Opinions in Connection with Digital Superintelligence. IAES International Journal of Artificial Intelligence. 2020; 9(2): 336-348.
9. Cochrane P. Exponential Technology and The Singularity. Ubiquity. 2014; 2014:1-9.
10. Cristianini N. On the current paradigm in artificial intelligence. AI Communications. 2014; 27(1): 37-43.
11. Al-Imam A, Al-Lami F. Machine Learning for Potent Dermatology Research and Practice. Journal of Dermatology and Dermatologic Surgery. 2020; 24(1): 1-4.
12. WHO EMRO | Early COVID-19 preparation saved lives in Iraq | Iraq-news | Iraq [Internet]. Emro.who.int. 2020 [cited 2 October 2020]. Available from: http://www.emro.who.int/irq/iraq-news/early-covid-19-preparation-saved-lives-in-iraq.html
13. Google Trends. 2020 [cited 2 October 2020]. Available from: https://trends.google.com/
14. Web Scraper - The #1 web scraping extension [Internet]. Webscraper.io. 2020 [cited 29 September 2020]. Available from: https://www.webscraper.io/
15. Møller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks. 1993; 6(4): 525-33.
16. Olson CF. Parallel algorithms for hierarchical clustering. Parallel Computing. 1995; 21(8): 1313-25.
17. OCEBM Levels of Evidence — Centre for Evidence-Based Medicine (CEBM), University of Oxford [Internet]. Cebm.ox.ac.uk. 2020 [cited 2 October 2020]. Available from: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence
18. Nature News. Engineered bat virus stirs debate over risky research [Internet]. Nature.com [cited 6 October 2020]. Available from: https://www.nature.com/news/engineered-bat-virus-stirs-debate-over-risky-research-1.18787
19. Menachery V, Yount B, Debbink K, Agnihothram S, Gralinski L, Plante J et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. Nature Medicine. 2015; 21(12): 1508-1513.
20. Lu D. The hunt to find the coronavirus pandemic's patient zero. New Scientist. 2020; 245(3276): 9.
21. Ball J. The zoo hypothesis. Icarus. 1973; 19(3): 347-349.
22. WHO EMRO | Iraq reports a noticeable increase in COVID-19 cases since February | Iraq-news | Iraq [Internet]. Emro.who.int. 2020 [cited 6 October 2020]. Available from: http://www.emro.who.int/irq/iraq-news/iraq-reports-noticeable-increase-in-covid-19-cases.html
23. Coronavirus: Iraq reports first two confirmed deaths as fears rise [Internet]. BBC News. 2020 [cited 6 October 2020]. Available from: https://www.bbc.com/news/world-middle-east-51751952
24. Iraq: Nationwide lockdown implemented March 22 /update 15 [Internet]. GardaWorld. 2020 [cited 6 October 2020]. Available from: https://www.garda.com/crisis24/news-alerts/325526/iraq-nationwide-lockdown-implemented-march-22-update-15
25. Iraq extends coronavirus lockdown until mid-April [Internet]. Al Arabiya English. 2020 [cited 6 October 2020]. Available from: https://english.alarabiya.net/en/News/middle-east/2020/03/26/Iraq-extends-coronavirus-lockdown-until-mid-April-
26. COVID-19! [Internet]. App Store. 2020 [cited 6 October 2020]. Available from: https://apps.apple.com/us/app/covid-19/id1504906590
27. Al-Imam A, Gorial FI, Al-shalchy A. A Novel Unusual Manifestation of CH-Alpha as Acute Metabolic Disturbances: Case Report and Big Data Analytics. Journal of the Faculty of Medicine Baghdad. 2020; 62(1,2): 41-47.
28. Al-Imam A, Sahai A, Al-Derzi AR, Al-Shalchy A, Abdullah F. "All Models Are Wrong, But Some Are Useful": On the Non-Bayesian Statistical Robustness of Hilton's Law. European Journal of Anatomy. 2020; 24(1): 75-78.
29. Al-Imam A. Optimizing Linear Models via Sinusoidal Transformation for Boosted Machine Learning in Medicine. Journal of the Faculty of Medicine Baghdad. 2019; 61(3,4): 128-136.
30. Al-Imam A, Abdul-Wahaab IT, Konuri VK, Sahai A. Reconciling artificial intelligence and non-Bayesian models for pterygomaxillary morphometrics. Folia Morphol (Warsz). 2021. doi: 10.5603/FM.a2020.0149. Epub ahead of print. PMID: 33438189.
31. Al-Imam, A. Inferential Analysis of Big Data in Real-Time: One Giant Leap for Spatiotemporal Digital Epidemiology in Dentistry. Odontostomatology Research Anatomy Learning & Implantology. 2019; 12(1): 1-14.
32. Mian A, Khan S. Coronavirus: the spread of misinformation. BMC Medicine. 2020; 18(1): 89. https://doi.org/10.1186/s12916-020-01556-3
33. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the Internet: Infodemiology Study. JMIR Public Health and Surveillance. 2020; 6(2): e18444.
34. Allen CM. Good news! You can be a mask skeptic and still wear one to prevent COVID-19 spread. Opinion [Internet] 2020 [cited 10 October 2020]. Available from: https://www.miamiherald.com/opinion/op-ed/article243720037.html
35. Branson-Potts H. A face mask is part of the 'scamdemic,' they say. But they'll be happy to sell you one [Internet] Los Angeles Time. 2020 [cited 10 October 2020]. Available from:

https://www.latimes.com/world-nation/story/2020-07-25/anti-mask-sales-coronavirus-texas-arizona

36. Bayram M, Springer S, Garvey C, Özdemir V. COVID-19 Digital Health Innovation Policy: A Portal to Alternative Futures in the Making. OMICS: A Journal of Integrative Biology. 2020; 24(8): 460-469.

37. Jha V, Bassi A, Arfin S, John O. An overview of mobile applications (apps) to support the coronavirus disease 2019 response in India. Indian Journal of Medical Research. 2020; 151(5): 468-473.

38. Radanliev P, De Roure D, Walton R, Van Kleek M, Montalvo RM, Santos O, et al. COVID-19 what have we learned? The rise of social machines and connected devices in pandemic management following the concepts of predictive, preventive and personalized medicine. EPMA Journal. 2020; 11(3): 311-332.

39. Swayamsiddha S, Mohanty C. Application of cognitive Internet of Medical Things for COVID-19 pandemic. Diabetes & Metabolic Syndrome. 2020; 14(5): 911-915.

40. 5. Yasaka T, Lehrich B, Sahyouni R. Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App. JMIR mHealth and uHealth. 2020; 8(4): e18936.

41. Nikolopoulos K, Punia S, Schäfers A, Tsinopoulos C, Vasilakis C. Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. European Journal of Operational Research. 2020.

مزايا شبكة الويب السطحي لجائحة كورونا في العراق

د. أحمد محمد لطفي الامام
أ. م. د. مارك موتيكا
د. هند جدوع الدوري

خلاصة البحث:
خلفية البحث: البيانات الخاصة بفيروس كورونا المستجد من البلدان النامية ليست دقيقة تمامًا، الأمر الذي يتطلب دمج بيانات الوبائيات الرقمية فيما يخص جائحة كورونا.
أهداف البحث: التوفيق بين النماذج غير البايزية والذكاء الاصطناعي فيما يتعلق بالبيانات الوبائية الرقمية والكلاسيكية (غير الرقمية) المرتبطة بجائحة كورونا في العراق.
طرائق البحث: تصميم الدراسة طولي، للفترة من 24 فبراير 2020 إلى 25 سبتمبر 2020. تم استرجاع البيانات من وزارة الصحة العراقية حول الحالات اليومية، والتعافي، والوفيات الناتجة عن فيروس كورونا المستجد، واستخلاص بيانات رقمية أضافية من على منصة Google Trends باستخدام خمسة مصطلحات بحث هي "Coronavirus" و "كورونا" و "COVID-19" و "كوفيد -19" و "لقاح كورونا". تمثل مصطلحات البحث "كورونا"، "كوفيد -19" و "لقاح كورونا"، الترجمات العربية لكل من "Coronavirus"، "COVID-19" و "COVID-19 Vaccine". قمنا بتنفيذ اختبارات متعددة المتغيرات والتعلم الآلي لفحص الدلالات المكانية والزمانية للجائحة في العراق، وتفسير السببية التي تؤثر على العراقيين للحصول على المعرفة الرقمية، عبر الويب، فيما يخص فيروس كورونا المستجد والجائحة التي تسبب بها.
النتائج: مثلت بغداد والسليمانية قيم إحصائية متطرفة فيما يتعلق بالحالات اليومية والتعافي، والوفيات اليومية، على التوالي. كشفت الاختبارات متعددة المتغيرات والشبكات العصبية الاصطناعية عن التأثير التنبؤي للوفيات وحالات التعافي والحالات اليومية على عمليات البحث على شبكة الأنترنت السطحي المتعلقة بمصطلحين للبحث عبر الشبكة العنكبوتية، "كورونا" و "فيروس كورونا". باستخدام التجميع الهرمي، حددنا مجاميع مميزة تشمل العاصمة العراقية وإقليم كردستان وجنوب العراق. كانت ثلاثة مصطلحات بحث هي الأكثر انتشارًا بين مستخدمي الأنترنت من العراقيين، وهي "كورونا"، "كوفيد -19"، و "Coronavirus". كانت جميع الارتباطات الاستثنائية ثنائية المتغير إيجابية ما عدا تلك التي تتضمن مصطلح البحث "لقاح كورونا". في حين كان سكان محافظة المثنى الأقل اهتمامًا، بين العراقيين، حول البيانات المتعلقة باللقاحات المحتملة لفيروس كورونا المستجد.
الاستنتاجات: كانت التحليلات منتصرة في التوفيق بين النماذج غير البايزية والتعلم الآلي، باستخدام صنفين من بيانات علم الأوبئة حول الجائحة في العراق. نرى أن الدراسة الحالية ثمينة لمتخذي القرار في وزارة الصحة العراقية.

الكلمات الدالة :الذكاء الاصطناعي؛ فيروس كورونا؛ كوفيد-19؛ علم الأوبئة الرقمي؛ علم الأوبئة؛ الإنترنت؛ فيروس كورونا المستجد؛ السارس-كوف-2؛ الويب السطحي.