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

DOI: https://doi.org/10.32007/jfacmedbagdad.624179

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

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

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

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", "كوفيد-19", "COVID-19", "فirus كورونا", and "لقاح كورونا". The search terms "كوفيد-19", "فirus كورونا", 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, "كوفيد-19" 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 "كوفيد-19", "فirus كورونا", 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 triumphs 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; coronavirus; 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
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", "كوفيد-19", "لقاح كوفيد-19", and "COVID-19 Vaccine" [13]. The three search terms "لاقاح كوفيد-19", "كوفيد-19", and "COVID-19 Vaccine" 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
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), “Coronavirus” (8.54, 0.418, 9.848, 122.711), and “ظافحة كورونا” (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 “ظافحة 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”, “ظافحة كورونا”, and “لقاح كورونا” 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), “كوفيد 19” (0.384, 210, p<0.001), “COVID-19” (0.217, 210, p<0.001), “Coronavirus” (0.278, 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.

Kendall’s Tau Bivariate Correlations
Significant bivariate correlations existed for "ظافحة كورونا" versus "Coronavirus" (Kendall’s tau_b=0.823, p-value<0.001), "لقاح كورونا" versus "كوفيد 19" (0.278, p<0.001), "لقاح كورونا" versus "كوفيد 19" (0.170, p=0.004), "لقاح كورونا" versus "Coronavirus" (0.180, p=0.001), "لقاح كورونا" versus "COVID-19" (0.178, p=0.007), and "لقاح كورونا" versus "كوفيد 19" (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 "لكونا" 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 "كوفيد 19", 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-11x^6 - 9E-06x^2 + 0.9724x^4 – 57015x^3 + 2E+09x^2 - 3E+13x + 2E+17, R^2=0.4325, Correlation Coefficient=0.6577), “كوفيد-19” (y = 6E-11x^6 - 2E-05x^3 + 1.7267x^4 - 101274x^3 + 3E+09x^2 - 6E+13x + 4E+17, R^2=0.2985, Correlation Coefficient=0.5464), and “كوفيد-2” (y = 2E-11x^6 - 6E-06x^3 + 0.6336x^2 - 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.

<table>
<thead>
<tr>
<th>Search Term</th>
<th>Equation of the Predictive Model **</th>
<th>Order of Polynomial Function</th>
<th>R^2 Score</th>
<th>Correlation Coefficient</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronavirus</td>
<td>y = 3E-11x^6 - 9E-06x^2 + 0.9724x^4 – 57015x^3 + 2E+09x^2 - 3E+13x + 2E+17</td>
<td>6^a</td>
<td>0.4325</td>
<td>0.6577</td>
<td>Moderate</td>
</tr>
<tr>
<td>كوفيد-19</td>
<td>y = 6E-11x^6 - 2E-05x^3 + 1.7267x^4 - 101274x^3 + 3E+09x^2 - 6E+13x + 4E+17</td>
<td>6^a</td>
<td>0.2985</td>
<td>0.5464</td>
<td>Moderate</td>
</tr>
<tr>
<td>كوفيد-2</td>
<td>y = 2E-11x^6 - 6E-06x^3 + 0.6336x^2 - 37154x^3 + 1E+09x^2 - 2E+13x + 2E+17</td>
<td>6^a</td>
<td>0.1784</td>
<td>0.4224</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

* 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", "كوفيد-19", and "كوفيد-2" 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.

REGION [IRAQ]
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 "كوفيد-19" (Pearson r=-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", "كوفيد-19", and "короноvа" 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 "كوفيد-19" (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, "كوفيد-19" (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).
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>
<thead>
<tr>
<th>Source</th>
<th>Dependent Variable</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>Coronavirus</td>
<td>419.562</td>
<td>3</td>
<td>139.854</td>
<td>5.432</td>
<td>.011</td>
<td>.538</td>
</tr>
<tr>
<td></td>
<td>كورونا</td>
<td>527.919</td>
<td>3</td>
<td>175.973</td>
<td>6.366</td>
<td>.006</td>
<td>.577</td>
</tr>
<tr>
<td></td>
<td>19-كوفرن</td>
<td>2.271</td>
<td>3</td>
<td>.757</td>
<td>.589</td>
<td>.633</td>
<td>.112</td>
</tr>
<tr>
<td>Recoveries</td>
<td>Coronavirus</td>
<td>14638.102</td>
<td>1</td>
<td>14638.102</td>
<td>568.567</td>
<td>.000</td>
<td>.976</td>
</tr>
<tr>
<td></td>
<td>كورونا</td>
<td>9094.906</td>
<td>1</td>
<td>9094.906</td>
<td>328.993</td>
<td>.000</td>
<td>.959</td>
</tr>
<tr>
<td></td>
<td>19-كوفرن</td>
<td>418.957</td>
<td>1</td>
<td>418.957</td>
<td>325.730</td>
<td>.000</td>
<td>.959</td>
</tr>
<tr>
<td>New Cases</td>
<td>Coronavirus</td>
<td>22.448</td>
<td>1</td>
<td>22.448</td>
<td>.872</td>
<td>.366</td>
<td>.059</td>
</tr>
<tr>
<td></td>
<td>كورونا</td>
<td>42.964</td>
<td>1</td>
<td>42.964</td>
<td>1.554</td>
<td>.233</td>
<td>.100</td>
</tr>
<tr>
<td></td>
<td>19-كوفرن</td>
<td>1.906</td>
<td>1</td>
<td>1.906</td>
<td>1.482</td>
<td>.244</td>
<td>.096</td>
</tr>
<tr>
<td>Deaths</td>
<td>Coronavirus</td>
<td>342.805</td>
<td>1</td>
<td>342.805</td>
<td>13.315</td>
<td>.003</td>
<td>.487</td>
</tr>
<tr>
<td></td>
<td>19-كوفرن</td>
<td>.001</td>
<td>1</td>
<td>.001</td>
<td>.000</td>
<td>.984</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>Coronavirus</td>
<td>360.438</td>
<td>14</td>
<td>25.746</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>كورونا</td>
<td>387.025</td>
<td>14</td>
<td>27.645</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19-كوفرن</td>
<td>18.007</td>
<td>14</td>
<td>1.286</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Coronavirus</td>
<td>55892.000</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>كورونا</td>
<td>24315.000</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19-كوفرن</td>
<td>1287.000</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>Coronavirus</td>
<td>780.000</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>كورونا</td>
<td>914.944</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19-كوفرن</td>
<td>20.278</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Table 4. Neural Network Information.

<table>
<thead>
<tr>
<th>Input Layer</th>
<th>Covariates</th>
<th>1 Recoveries</th>
<th>2 New Cases</th>
<th>3 Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Units</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rescaling Method for Covariates Standardized</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hidden Layer(s)</td>
<td>1</td>
<td>Number of Hidden Layers</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Number of Units in Hidden Layer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activation Function</td>
<td>Hyperbolic tangent</td>
<td>1 Coronavirus</td>
<td>2 كورونا</td>
<td></td>
</tr>
<tr>
<td>Output Layer</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Number of Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rescaling Method for Scale Standardized</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependents</td>
<td></td>
<td>Identity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error Function</td>
<td>Sum of Squares</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Excluding the bias unit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

J Fac Med Baghdad 122 Vol.62 No.4, 2020
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, respectively.

Discussion:
Three search terms were most prevalent among Iraqi users of the surface web, including "Coronavirus", "كورونا" and "Coronavirus-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 "افةك 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

Figure 4. Supervised Machine Learning: Neural Network Analysis.

Figure 5. Unsupervised Machine Learning: Dendrogram of Hierarchical Cluster Analysis.
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, covariates, 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 disforms 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
7. Plebanski M, Apostolopoulos V. Moderna’s COVID vaccine reports 95% efficacy. It means we
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/oceb-levels-of-evidence
35. Branson-Potts H. A face mask is part of the 'scamademic,' they say. But they'll be happy to sell you one [Internet] Los Angeles Time. 2020 [cited 10 October 2020]. Available from:
Surface Web Merits for SARS-CoV-2 Pandemic in Iraq.

Ahmed Al-Imam


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

Ahmed Al-Imam

Mazaya: The Surface Web Merits for SARS-CoV-2 Pandemic in Iraq

Ahmed Al-Imam

J Fac Med Baghdad 127 Vol.62 No.4, 2020

Abstract:

Background: The data on the latest SARS-CoV-2 virus disease cases in developing countries is not entirely accurate, which requires the integration of digital epidemiological data to improve the impact of the epidemic.

Aim: To compare the Bayesian models with artificial intelligence in terms of the digital epidemiological and classic (non-digital) data of the SARS-CoV-2 epidemic in Iraq.

Method: The study was designed long, from February 24, 2020 to September 25, 2020. The data were extracted from the Ministry of Health in Iraq regarding daily cases, recoveries, and deaths from the SARS-CoV-2 virus, and additional digital data were used through Google Trends. Five keywords were used in the search: "Coronavirus" and "كوفيد-19" and "COVID-19" and "COVID-19 Vaccine" and "قاح كورونا". We ran multiple tests and artificial intelligence to analyze the spatial and temporal impacts of the epidemic in Iraq, and we interpreted the factors that affect Iraqis to access digital information on the Internet concerning SARS-CoV-2 virus and the epidemic that has been caused by it.

Results: Baghdad and Salaheddin had extreme statistical values in daily cases, recoveries, and deaths in turn. Multiple comparisons of variables and artificial neural networks showed the predictive effect of deaths and recoveries and daily cases on internet searches related to the two keywords of the search through the web, "Coronavirus" and "فيروس كورونا". Using the hierarchical grouping, we determined distinctive groups that included the capital and Kurdistan and southern Iraq. Three keywords were the most popular among Internet users in Iraq, "Coronavirus", "كوفيد-19" and "Coronavirus". All the relationships were positive except those that included the search "لقاح كورونا". In contrast, the residents of the Muthanna province were less interested in the data concerning the potential vaccines of the SARS-CoV-2 virus.

Conclusion: It is clear that the analyses were successful in reconciling the Bayesian models with artificial intelligence, using two categories of epidemiological data related to the SARS-CoV-2 epidemic in Iraq. We believe that the current study is valuable for the health authorities in Iraq.

Keywords: Artificial intelligence; SARS-CoV-2 virus; "Coronavirus"; Epidemiological digital science; Epidemiological; Internet; SARS-CoV-2 virus; Surface Web.