International Journal on Data Science and Technology
2022; 8(2): 36-42
http://www.sciencepublishinggroup.com/j/ijdst
doi: 10.11648/j.ijdst.20220802.11
ISSN: 2472-2200 (Print); ISSN: 2472-2235 (Online)

Using Data Mining Techniques for COVID-19: A Systematic Review

Sanjib Ghosh[1, 3], Lipon Chandra Das[2, 4, *]

[1]Department of Statistics, University of Chittagong, Chittagong, Bangladesh

[2]Department of Mathematics, University of Chittagong, Chittagong, Bangladesh

[3]School of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou, China

[4]Climate Modeling Laboratory, School of Mathematics, Shandong University, Jinan, China

Email address:

liponbgc@gmail.com (L. C. Das)

[*]Corresponding author

To cite this article:

Sanjib Ghosh, Lipon Chandra Das. Using Data Mining Techniques for COVID-19: A Systematic Review. *International Journal on Data Science and Technology*. Vol. 8, No. 2, 2022, pp. 36-42. doi: 10.11648/j.ijdst.20220802.11

Received: April 2, 2022; **Accepted**: June 1, 2022; **Published**: June 16, 2022

Abstract: The primary goal of this survey is to determine the most widely used data mining approaches and knowledge gaps from published publications. The novel coronavirus pneumonia, namely *COVID-19*, has become a global public health problem. Since the threat of pandemics has raised public health concerns, researchers to uncover hidden knowledge have used data extraction techniques. Web of Science, Scopus, and PubMed databases were used to conduct systematic research. Then, to choose good papers, all retrieved publications were reviewed in a stepwise procedure using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist. All of the data were examined and summarized using a few different classifications. Out of 300 citations, 50 papers were eligible through a systematic review. The review results showed that the most favorite DM belonged to Natural language processing (22%), and the most commonly proposed approach was revealing disease characteristics (22%). Regarding diseases, the most addressed disease was *COVID-19*. The studies predominately apply supervised learning techniques (90%). We found infectious disease (36%) to be the most frequent, closely followed by epidemiology discipline concerning healthcare scopes. The most common software used in the studies was SPSS (22%) and R (20%). Our results indicate that there is a significant relationship between air pollution and *COVID-19* infection, which could partially explain the effect of national lockdown and provide implications for the control and prevention of this novel disease. The results revealed valuable research conducted by employing the capabilities of knowledge discovery methods to understand the unknown dimensions of diseases in pandemics. However, most research will need in terms of treatment and disease control.

Keywords: Public Health, COVID-19, SARS-CoV-2, Machine Learning, Meta-Analyses, Data Mining

1. Introduction

The possibility of pandemics has been a cause of concern for the medical profession throughout history. The prospect of major infectious illnesses spreading worldwide before anybody notices it is a contentious issue. In the past, the apparent frequency of Severe Acute Respiratory Syndrome (SARS) and other forms of influenza showed how a pandemic disease might damage a country's health system [1, 2]. Coronavirus illness (COVID-19) is the most recent pandemic disease series with worldwide ramifications. COVID-19, also known as the new Coronavirus (2019-nCoV), is a coronavirus 2 (SARS-CoV-2)-related viral illness that emerged on December 8, 2019 in Wuhan, China [3, 4]. The globe confronts major hurdles in controlling this epidemic since a novel coronavirus (nCoV) is a new strain of the coronavirus family that has never been detected previously. During epidemics, clinical professionals have attempted to develop treatments and vaccines, but scientists working in the fields of data science and technology have attempted to identify the infectious agent and assist in its management using information-based approaches [5, 6].

Over the past decades, valuable studies have been published regarding pandemics and data mining (DM) techniques [7]. These studies were carried out to understand better, control, and manage pandemics through various data mining methods. Due to the importance of dealing with the COVID-19 pandemic, it is essential to survey the most popular and efficient data mining methods that could significantly impact selecting the most effective techniques in pandemic studies. It can help us uncover the unknown nature of the new pandemic and future pandemics. This study aims to collect, synthesize, and evaluate existing publications to make monitoring and analysis of published studies on pandemics and data mining approaches easier. The specific research questions of this review are:

1. To determine how many studies have been published in recent years and months regarding previous pandemics and the COVID-19 outbreak.
2. To represent an overview of published studies and their characteristics.
3. To investigate published studies regarding data mining techniques.
4. To identify the source of data.
5. To determine the most popular DM techniques in frequency and clinical domains.
6. To determine the source of data.

COVID-19, a contagious disease caused by the SARS-COV-2 virus, necessitates extraordinary, high-intensity, high-potential responses in more than 200 countries worldwide. In the first four months of the epidemic, infected people ranged from 2 to 20 million, with at least 200,000 deaths. To combat the rapid spread of the COVID-19 illness among humans, all governments worldwide took drastic measures, including the quarantine of hundreds of millions of citizens [18]. Despite the difficulties in identifying positive and negative COVID-19 persons based on the many COVID-19 symptoms, all of these attempts are limited. Therefore, tests to detect the SARS–CoV-2 virus are believed to be critical to recognize the positive cases of this infection to limit the [19]. Radiology and imaging are some of the most beneficial and critical modalities used for diagnosis COVID-19 stage and hazards on the patient's lungs, specifically by chest CT scan [8]. Early diagnosis of COVID-19 is vital to minimize human-to-human transmission and patient care. Recently, the separation and quarantine of healthy people from the infected or per-sons who suspect that they are carrying the virus is the most effective technique for preventing the spread of COVID-19 [9]. The use of machine-learning techniques has revealed new insights into COVID-19 diagnoses, such as whether a lung computed tomography (CT) scan should be used as a first-line screening test or as an alternative to the real-time inverse transcriptase-polymerase chain reaction (RT–PCR) and the differences between COVID-19 pneumonia and other viral pneumonia using a CT scan of the lungs [10].

2. Methodology

This investigation was done using the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) checklist [14]. Then, to classify the primary characteristics of studies, a synthesis of eligible publications was undertaken based on the main characteristics. From December 8, 2019 to October 16, 2020 a systematic search of the scientific database, Web of Science, Scopus, and Pub Med databases was conducted using the keywords "data mining", "prediction model", "data mining techniques", "data mining methods", "pandemics", "pandemic", "COVID-19", "SARS-CoV-2", and "corona-virus disease." The keywords in each database were used to create Boolean search techniques.

The following criteria were used to determine whether or not an article should be included:

1) Only papers related to the application of data extraction techniques or knowledge discovery methods were included in this study;
2) Only articles connected to pandemic illnesses like COVID-19 were included. Because of the wide range of approaches utilized in this sector, these methods were chosen based on the findings of Patel and Patel [15]. If an article matched the following conditions, it was ruled out:
   i). The article's title, summary, or full text did not include pandemics or COVID-19 diseases.
   ii). Book chapters, letters to editors, short briefs, reports, commentaries, technical reports, reviews, or meta-analyses were not considered.
   iii). Non-English publications.
   iv). The complete text was not accessible.

Through the online interface of scientific websites, 300 articles were obtained from scientific databases (Web of Science, Scopus, and PubMed). For screening articles, inclusion and exclusion criteria were established. All titles and abstracts of retrieved papers were reviewed in the first step to identify relevant research. Three reviewers (MT, SS, and SR) examined all the titles and abstracts for relevant publications. Another reviewer (MG) looked at a random selection of studies. The Joanna Briggs Institute (JBI) checklist, which provides comprehensive checklists for the appraisal and assessment of most types of the research [16, 17], and was used to assess the quality of the individual publications. We used this checklist since our review includes a variety of research study-related decisions. Two reviewers made decisions on study eligibility and quality; any disagreements were resolved by discussion. The flow of screening articles based on the [17] PRISMA method illustrates in Figure 1.

Finally, 50 studies remained as eligible articles. Some classifications were assumed to classify and analyze inclusive studies. All eligible papers that met our inclusion criteria included 47 journal papers and three conference papers. As it is apparent, the majority of studies were published in 2020.

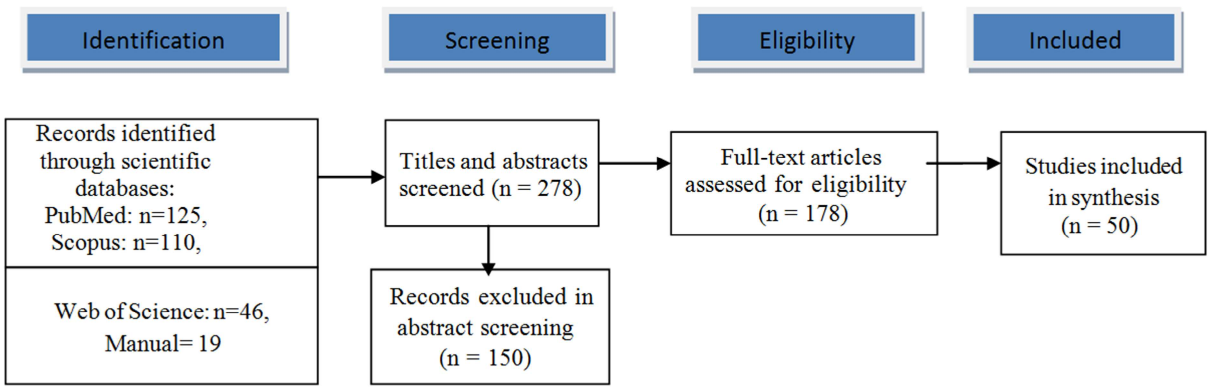

Figure 1. *The PRISMA diagram for the identification, screening and eligibility.*

Table 1 summarizes only 20 articles out of 50 included under predefined categories. All articles are summed in the word cloud in Figure 2 to see the frequency of terms that appeared more often in reviewed publications.

Table 1. *The characteristics of 20 reviewed articles.*

Author	Main approaches	Clinical scope	The applied method of data mining	Software (Environment)	Data source
Abd-Alrazaq A et al. [30]	Infoveillance	Social behavior	Text mining	Python	Twitter
Ahamad MM [19]	Disease characteristics	Infectious disease	Decision Tree, Random Forest, gradient boosting Machine, SVM	SPSS	Github repository
Ren X et al.	Treatment	Pharmacology	Association rule mining method, and association knowledge network	R	Traditional Chinese medicine system pharmacology database
Zhang Y et al. [31]	Infoveillance	Psychology	Time series, NLP, and deep learning	Python	Weibo social network
Sudirman ID Nugraha DY [59]	Risk factors	Infectious	Naive Bayes method	Rapid Miner	Ministry of Public Health Thailand
Huang C et al. [20]	Disease characteristics	Infectious disease	Text mining	Python	Sina Weibo social network
Han X et al. [32]	Infoveillance	Infectious disease	Time series, Random forest, Spatial Distribution	Python	Sina Weibo social network
Ibrahim et al. [61]	Tracing transmission	Epidemiology	ANN	not mentioned	CDC
Foieni F et al.[22]	Disease characteristics	Respiratory medicine	Multivariate Regression	SPSS	WHO
Zhao ZR et al. [46]	Patient monitoring and follow- up	Respiratory medicine Regression model	SPSS	COVID-19 PUI registry	Respiratory medicine Regression model
Fan Q et al. [60]	Risk factors	Cardiology	Logistic regression	SPSS	Wuhan Tongji hospital
Lei MT et al. [62]	Tracing transmission	Epidemiology	CART, Linear regression	SPSS	Macao Meteorological and Geo- physical Bureau
Dong YL et al. [42]	Patient monitoring and follow-up	up Infectious disease	Logistic regression	SPSS	Wuhan union hospital
Roland LT et al. [26]	Disease characteristics	Respiratory medicine	Logistic regression	SPSS	San Francisco (USF) institutional
Zhou YW et al.[51]	Early diagnosis	Infectious disease	Logistic regression, Nomograms	R	47 locations in Sichuan province
Li S et al. [54]	Early diagnosis	Psychology	Text mining	SPSS	Weibo posts
Ayyoubzadeh SM et al. [34]	Infoveillance	Epidemiology scope Linear regression and long short	term memory (LSTM) models	Python	Google data
Qiang X et al. [50]	Active case prediction	Infectious disease	Random forest (RF) method	R	China national genomics data center
Liu. Q et al. [27]	Disease characteristics	Infectious disease	Logistic regression	SPSS	Union Hospital, Tongji medical
KostkovaP et al. [41]	Outbreak prediction	Public health	Text mining	Not mentioned	Twitter
Kostoff RN [35]	Infoveillance	Informatics	Text mining	Not mentioned	Medical literature
Szomszo M et al. [36]	Infoveillance	Informatics	Text mining, linked resource	Not mentioned	Twitter

According to reviewed studies, we can classify all eligible articles in this review into eight categories based on their clinical discipline. The identified clinical and health disciplines with their distribution and their frequency are described in Figure 2.

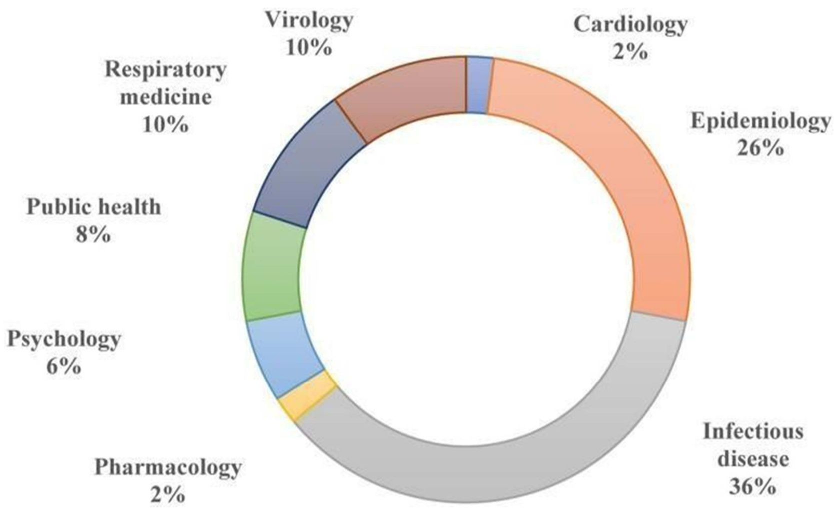

Figure 2. *The frequency of main health disciplines in reviewed articles.*

According to the chart, it is evident that the greatest demand was related to infectious disease with 18 papers (36%). Next, epidemiology is the second largest discipline of included studies with 13 studies (26%). This analysis can be very helpful in determining gaps in the health domains.

Since the main objective of this study was to deter- mine to what extent data mining techniques employed to fight pandemics, the frequency of applied methods was investigated in this section according to a study conducted by Patel and Patel [15]. Table 2 showed an overview of the distribution of applied data mining methods in reviewed articles. The analysis showed that all of the applied methods classified into 14 main categories. It is apparent that the most favorite method was employed in reviewed articles belonged to Natural language processing (NLP) techniques (22%). While logistic regression analysis with 20% of studies was in the second rank to determine the association of the independent variables with a dichotomous dependent variable [67].

Table 2. *Frequency of data mining techniques in reviewed studies.*

DM techniques	Frequency	Percentage	Studies
NLP techniques	11	22.00%	[20–30]
Logistic regression	10	20.00%	[31–40]
Time series	7	14.00%	[20, 41–46]
Random forest	7	14.00%	[47-51, 45, 42]
Regression models	7	12.00%	[52-55, 40, 49, 39]
Decision tree	6	12.00%	[51, 48, 56–58, 39]
ANN	5	10.00%	[52, 59, 60, 21]
Naive Bayes	3	6.00%	[61–63]
SVM	2	4.00%	[49, 51]
Association rule mining	2	4.00%	[65, 58, 66]
Clustering	2	4.00%	[34, 30]
Apriority algorithm	1	2.00%	[64]
Genetic algorithm	1	2.00%	[55]
Fuzzy algorithm	1	2.00%	[41]

3. Discussion

The primary goal of this study was to compile a list of papers on the use of data based DM techniques in pandemics. As a result, 50 publications were chosen and examined among 300 investigations. For the research that was included, a variety of data sources were employed. Most research was carried out in China because most pandemics started in this nation.

According to a qualitative study, researchers chose to employ supervised approaches like regression to create prediction models for a better knowledge of unknown pandemics. All of these strategies have been successfully applied in diverse fields of medicine [11]. In addition, classification methodologies were used more than expected in the studies. By selecting the best method to implement exact prediction models, researchers can discover some biomarkers in unknown diseases that can enable them to predict essential outcomes [12, 13]. Therefore, developing prediction models can help physicians and help health policymakers and societies.

The data revealed that preventing infectious illness transmission is the most critical priority in pandemic disease [68]. In general, the nature of a new illness during a pandemic is unclear, and scientists are concerned about recognizing the hallmarks of a new disease. That is why the majority of research is focused on identifying disease features. During this epidemic, scientists should prioritize diagnosis above other duties [69]. The second primary concern with pandemic infections is how they propagate. As a result, about 10% of studies have been devoted to forecasting the disease's occurrence.

However, because of the various methodologies used, the sample size of datasets varies greatly. The findings revealed that most of the research employed a limited number of data sets and a variety of data sources. The use of massive datasets can increase the model's precision and robustness of results [70], which can help scientists better combat this emerging illness. As a result, researchers should employ massive datasets in their studies, even if they are conducted worldwide, to make better diagnostic and treatment recommendations. We ran upon certain limits in our research. As a result, some research may be overlooked when this article is published.

4. Conclusion

This review could help scientists to reach published researches regarding DM techniques and fierce pandemics easier. We present a review of the *COVID-19* literature and identify current research hotspots and priorities. Our findings can assist the research community in identifying and prioritizing research needs, as well as identifying prominent *COVID-19* researchers, institutions, nations, and publications. Our study provided a systematic review of an exhaustive overview of integrated AI-based DM and machine learning algorithms with the CoV family. The data mining approaches used in worldwide pandemics were investigated in this study. However, to avoid and anticipate the *COVID-19* outbreak, most of these strategies were created in the current context. According to our survey, we found out that the foremost objective of DM applications is related to disease characteristics. We think that our study may be applied to additional eHealth-related publications to provide physicians, administrators, and policymakers with a comprehensive perspective of the literature and the ability to categorize distinct subjects of current research for further investigation.

References

[1] Jain V, Duse A, Bausch DG. Planning for large epidemics and pandemics: challenges from a policy perspective. Curr Opin Infect Dis. 2018; 31 (4): 316–24.

[2] Cook AH, Cohen DB. Pandemic Disease: A Past and Future Challenge to Governance in the United States. Rev Policy Res. 2008; 25 (5): 449–71.

[3] Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. Features, evaluation, and treatment coronavirus (COVID-19). StatPearls [Internet]. StatPearls Publishing 2020.

[4] Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. Estimation of the reproductive number of novelcoronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. Int J Infect Dis. 2020; 93: 201–4.

[5] Atif I, Cawood FT, Mahboob MA. The Role of Digital Technologies that Could Be Applied for Prescreening in the Mining Industry During the COVID-19 Pandemic. Transactions of the IndianNational Academy of Engineering. 2020: 1–12.

[6] Gholamzadeh M, Abtahi H, Safdari R. Suggesting a framework for preparedness against thepandemic outbreak based on medical informatics solutions: a thematic analysis. The International Jour- nal of health planningand management. 2021; n/a (n/a).

[7] Gulyaeva M, Huettmann F, Shestopalov A, Okamatsu M, Matsuno K, Chu D-H, et al. Data mining and model-predicting a global dis- ease reservoir for low-pathogenic Avian Influenza in the wider pacific rim using big data sets.

[8] Day M (2020) COVID-19: identifying and isolating asymptomatic peo- ple helped eliminatevirus in Italian village. BMJ 368: 135.

[9] Deng X et al (2020) A classification–detection approach of COVID-19 based on chest X- rayand CT by using keras pre-trained deep learning models. Comput Model Eng Sci 125 (2): 579–596.

[10] Kassani SH et al (2020) Automatic detection of coronavirus disease (COVID-19) in X-ray andCT images: a machine learning-based approach 10 (4): 1–18.

[11] Asur S, Huberman BA, editors. Predicting the future with social media. 2010 IEEE/WIC/ACMinternational conference on web intelligence and intelligent agent technology; 2010: IEEE.

[12] Zhang Y, Guo SL, Han LN, Li TL. Application and Exploration of Big Data Mining in ClinicalMedicine. Chin Med J. 2016; 129 (6): 731–8.

[13] Alanazi HO, Abdullah AH, Qureshi KN, Ismail AS. Accurate and dynamic predictive modelfor better prediction in medicine and healthcare. Irish Journal of Medical Science (1971 -).2018; 187 (2): 501–13.

[14] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ (Clinical research ed). 2009; 339: b2535.

[15] Patel S, Patel H. Survey of Data Mining Techniques used in Healthcare Domain. International Journal of Information Sciences and Techniques. 2016; 6 (1/2): 53–60.

[16] Institute JB. The Joanna Briggs Institute Critical Appraisal Tools. University of Adelaide, South Australia. 2017. https://jbi.global/critical-appraisal-tools. Accessed 5 Mar 2021.

[17] Hannes K, Lockwood C, Pearson A. A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. Qual Health Res. 2010; 20 (12): 1736–43.

[18] Alimadadi A et al (2020) Artificial intelligence and machine learning to fight COVID-19.

[19] American Physiological Society, Bethesda Brinati D et al (2020) Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. J Med Syst 44: 135.

[20] Huang C, Xu X, Cai Y, Ge Q, Zeng G, Li X, et al. Mining the characteristics of COVID-19 patients in China: Analysis of social media posts. J Med Internet Res. 2020; 22 (5).

[21] Maram B, Padmapriya G, Satish AR. A framework for performance analysis on machine learning algorithms using COVID-19 dataset. Adv Math: Sci J. 2020; 9 (10): 8207-15.

[22] Foieni F, Sala G, Mognarelli JG, Suigo G, Zampini D, Pistoia M, et al. Derivation and validation of the clinical prediction model for COVID-19. Intern Emerg Med. 2020.

[23] Ma XX, Li A, Jiao MF, Shi QM, An XC, Feng YH, et al. Character- istic of 523 COVID-19 inHenan Province and a Death Prediction Model. Frontiers in Public Health. 2020; 8.

[24] Luo Y, Mao LY, Yuan X, Xue Y, Lin Q, Tang GX, et al. Prediction Model Based on the Combination of Cytokines and Lymphocyte Subsets for Prognosis of SARS-CoV-2 Infection. J Clin Immunol. 2020; 40 (7): 960–9.

[25] Ciucurel C, Iconaru EI. An Epidemiological Study on the Prev- alence of the Clinical Featuresof SARS-CoV-2 Infection in Romanian People. Int J Environ Res Public Health. 2020; 17 (14).

[26] Roland LT, Gurrola JG, Loftus PA, Cheung SW, Chang JLL. Smell and taste symptom- based predictive model for COVID-19 diagnosis. International Forum of Allergy & Rhinology. 2020; 10 (7): 832–8.

[27] Liu Q, Song NC, Zheng ZK, Li JS, Li SK. Laboratory findings and a combined multifactorialapproach to predict death in critically ill patients with COVID-19: a retrospective study. Epidemiology and Infection. 2020; 148.

[28] Li D, Chaudhary H, Zhang Z. Modeling Spatiotemporal Pattern of Depressive Symptoms Caused by COVID-19 Using Social Media Data Mining. Int J Environ Res Public Health. 2020; 17 (14).

[29] Sarker A, Lakamana S, Hogg-Bremer W, Xie A, Al-Garadi MA, Yang YC. Self- reported COVID-19 symptoms on Twitter: an analysis and a research resource. J Am Med InformAssoc. 2020; 27 (8): 1310–5.

[30] Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters Duringthe COVID-19 Pandemic: Infoveil- lance Study. J Med Internet Res. 2020; 22 (4): e19016. https://doi. org/10.2196/19016.

[31] Zhang Y, Cheng J, Yang Y, Li H, Zheng X, Chen X, et al. COVID-19 public opinion and emotionmonitoring system based on time series thermal new word mining. Comput Mater Continua. 2020; 64 (3): 1415–34.

[32] Han X, Wang J, Zhang M, Wang X. Using social media to mine and analyze public opinionrelated to COVID-19 in China. Int J Environ Res Public Health. 2020; 17 (8). https://doi.org/10.3390/ ijerph17082788.

[33] Qin L, Sun Q, Wang Y, Wu KF, Chen M, Shia BC, et al. Predic- tion of number of cases of 2019 novel coronavirus (COVID-19) using social media search index. Int J Environ ResPublic Health. 2020; 17 (7).

[34] Ayyoubzadeh SM, Ayyoubzadeh SM. Predicting COVID-19 Incidence Through Analysis of GoogleTrends Data in Iran: Data Mining and Deep Learning Pilot Study. JMIR Public Health Surveill. 2020; 6 (2): e18828.

[35] Kostoff RN. Literature-related discovery: potential treatments and preventatives for SARS. Technol Forecast Soc Chang. 2011; 78 (7): 1164–73.

[36] Szomszor M, Kostkova P, St Louis C, editors. Twitter informatics: tracking and understandingpublic reaction during the 2009 swine flu pandemic. 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology; 2011: IEEE.

[37] Wahbeh A, Nasralah T, Al-Ramahi M, El-Gayar O. Mining Physi- cians' Opinions on SocialMedia to Obtain Insights Into COVID-19: Mixed Methods Analysis. JMIR Public HealthSurveill. 2020; 6 (2): e19276.

[38] Chintalapudi N, Battineni G, Amenta F. COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach. Journal ofMicrobiology, Immunology and Infection. 2020; InPress.

[39] Khan MA, Abbas S, Khan KM, Al Ghamdi MA, Rehman A. Intelligent Forecasting Model of COVID-19 Novel Coronavirus Outbreak Empowered with Deep Extreme Learning Machine. Cmc-Computers Materials & Continua. 2020; 64 (3): 1329–42.

[40] Kargarfard F, Sami A, Hemmatzadeh F, Ebrahimie E. Identifying mutation positions in all segments of influenza genome enables better differentiation between pandemic and seasonalstrains. Gene. 2019; 697: 78–85.

[41] Kostkova P, Szomszor M, St. Louis C. Swineflu: The use of twitter as an early warning and riskcommunication tool in the 2009 swine flu pandemic. ACM Transactions on Management Information Systems (TMIS). 2014; 5 (2): 1–25.

[42] Dong YL, Zhou HF, Li MY, Zhang ZL, Guo WN, Yu T, et al. A novel simple scoring model forpredicting severity of patients with SARS-CoV-2 infection. Transboundary and Emerging Dis. 2020.

[43] Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severityof coronavirus disease 2019: a model-based analysis. The Lancet Infectious Dis. 2020; InPress.

[44] Ghosal S, Sinha B, Majumder M, Misra A. Estimation of effects of nationwide lockdown for containing coronavi- rus infection on worsening of glycosylated haemoglobin and increase in diabetes-related complications: A simulation model using multivariate regressionanalysis. Diabetes Metab Syndr. 2020; 14 (4): 319–23.

[45] Cheng FY, Joshi H, Tandon P, Freeman R, Reich DL, Mazumdar M, et al. Using Machine Learning to Predict ICU Transfer in Hospitalized COVID-19 Patients. Journal of ClinicalMedicine. 2020; 9 (6).

[46] Zhao ZR, Chen AN, Hou W, Graham JM, Li HF, Richman PS, et al. Prediction model andrisk scores of ICU admission and mortality in COVID-19. Plos One. 2020; 15 (7).

[47] Alzahrani SI, Aljamaan IA, Al-Fakih EA. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA pre- diction model under current public health interventions. J Infect Public Health. 2020; 13 (7): 914–9.

[48] Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modeling and forecasting of the COVID-19 outbreak. PLoS ONE. 2020; 15 (3).

[49] Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R. COVID-19 Pandemic Prediction forHungary; A Hybrid Machine Learn- ing Approach. Mathematics. 2020; 8 (6).

[50] Qiang XL, Xu P, Fang G, Liu WB, Kou Z. Using the spike pro- tein feature to predict infectionrisk and monitor the evolution- ary dynamic of coronavirus. Infect Dis Poverty. 2020; 9 (1): 33.

[51] Zhou YW, He YQ, Yang H, Yu H, Wang T, Chen Z, et al. Devel- opment and validation a nomogram for predicting the risk of severe COVID-19: A multi-center study in Sichuan, China. Plos One. 2020; 15 (5). 10.1371/journal.pone.0233328.

[52] Yan L, Zhang HT, Goncalves J, Xiao Y, Wang ML, Guo YQ et al. An interpretable mortalityprediction model for COVID-19 patients. Nature Machine Intelligence. 2020; 2 (5): 283.

[53] Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, et al. Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coronavirus Clinical Severity. Computers, Mate- rials & Continua. 2020; 63 (1).

[54] Li S, Wang Y, Xue J, Zhao N, Zhu T. The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users. Int J Environ Res Public Health. 2020; 17 (6): 2032.

[55] Martin-Rodriguez F, Sanz-Garcia A, Lopez-Izquierdo R, Benito JFD, Martin-Conty JL, Villamor MAC, et al. Predicting Health Care Workers' Tolerance of Personal ProtectiveEquipment: An Observational Simulation Study. Clin Simul Nurs. 2020; 47: 65–72.

[56] Ketu S, Mishra PK. Enhanced Gaussian process regression-based forecasting model forCOVID-19 outbreak and signifi- cance of IoT for its detection. Appl Intell. 2020.

[57] Moftakhar L. The Exponentially Increasing Rate of Patients Infected with COVID-19 in Iran. Archives of Iranian medicine. 2020; 23 (4): 235–8.

[58] Yongjian Z, Jingu X, Fengming H, Liqing C. Association between short-term exposure to air pollution and COVID-19 infection: Evi- dence from China. Sci Total Environ. 2020: 138704.

[59] Sudirman ID, Nugraha DY. Naive Bayes classifier for predicting the factors that influence death due to COVID-19 in China. J Theor Appl Inf Technol. 2020; 98 (10): 1686–96.

[60] Sudirman ID, Aryanto R, Mulyani. Optimizing decision tree criteria for predicting COVID-19 mortality in South Korea dataset. J Theor Appl Inf Technol. 2020; 98 (15): 2889–900.

[61] Fan Q, Zhu HL, Zhao JX, Zhuang LF, Zhang H, Xie HY. Risk factors for myocardial injury in patients with coronavirus dis- ease et al 2019 in China Esc Heart Failure 2020.

[62] Ibrahim S, Kamaruddin SA, Sabri N, Samah KA, Noordin M, Shari A. The influences of global geographical climate towards COVID-19 spread and death. Int J Adv Trends Comput Sci Eng. 2020; 9 (1.4 Special Issue): 612–7.

[63] Lei MT, Monjardino J, Mendes L, Goncalves D, Ferreira F. Statistical Forecast of Pollution Episodes in Macao during National Holidayand COVID-19. Int J Environ Res Public Health. 2020; 17 (14).

[64] Ren X, Shao XX, Li XX, Jia XH, Song T, Zhou WY, et al. Identi- fying potential treatments of COVID-19 from Traditional Chinese Medicine (TCM) by using a data- driven approach. J Ethnophar- macol. 2020; 258.

[65] Neuraz A, Lerner I, Digan W, Paris N, Tsopra R, Rogier A, et al. Natural Language Processing for Rapid Response to Emergent Dis- eases: Case Study of Calcium Channel Blockers and Hypertension in the COVID-19 Pandemic. J Med Internet Res. 2020; 22 (8): e20773.

[66] Masand VH, Rastija V, Patil MK, Gandhi A, Chapolikar A. Extending the identification of structural features responsible for anti-SARS-CoV activity of peptide-type compounds using QSAR modelling. SAR QSAR Environ Res. 2020; 31 (9): 643–54.

[67] Kargarfard F, Sami A, Ebrahimie E. Knowledge discovery and sequence-based prediction of pandemic influenza using an integrated classification and association rule mining (CBA) algorithm. J Biomed Inform. 2015; 57: 181–8.

[68] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction modelsfor diagnosis and prognosis of COVID-19 infection: systematic review and critical appraisal. BMJ (Clinical research ed). 2020; 369: m1328-m. https://doi.org/10. 1136/bmj.m1328.

[69] Adhikari SP, Meng S, Wu Y-J, Mao Y-P, Ye R-X, Wang Q-Z, et al. Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. Infect Dis Poverty. 2020; 9 (1): 29.

[70] Kelly-Cirino CD, Nkengasong J, Kettler H, Tongio I, Gay-Andrieu F, Escadafal C, et al. Importance of diagnostics in epidemic and pandemic preparedness. BMJ global health. 2019; 4 (Suppl 2): e001179-e.