Journal of Positive Sciences (JPS) **eISSN: 2582-9351, https://positive-sciences.com** **JPS. 2021 Aug; 1(5): 15-18**

Article info **Research Article**
Received on: 23.07.2021
Accepted on: 29.08.2021
Published on: 31.08.2021
doi: https://doi.org/10.52688/ASP84514

Data mining methods for covid-19 influences detection on technological tools and instrumentations performance

Mohammed Hamid Abood [1, *]

[1] Al kunooze University College, Iraq

[*] mohammed.h@kunoozu.edu.iq

ABSTRACT

The primary aim of the survey is to examine the published papers to identify the most popular techniques and knowledge gaps for data mining. Since the danger of pandemics heightened public health worries, the researchers used data mining methods to uncover buried information. For systematic searches, websites for Science, Scopus, and PubMed databases were chosen. Then all papers found were evaluated in the process according to the Systematic Review and Meta-Analyses checklist Preferred Reporting Items for the selection of suitable publications. All findings were evaluated and presented on the basis of certain classifications. Of the 335 citations, 50 publications were selected by a scope evaluation as qualifying papers. The findings of the study indicate that the most popular DM was natural language processing (22%), with the most frequent method disclosing illness features (22 percent). Concerning illnesses, COVID-19 was the disease most addressed. The findings indicate that supervised learning methods predominate (90 percent). With respect to the healthcare sector, we discovered that infectious illness (36%) is the most common ailment, followed closely by disciplinary epidemiology. SPSS (22 percent) and R were the most prevalent software in the research (20 percent). The findings showed that some important study was carried out using the capacities of techniques of finding information to comprehend the unknown aspects of pandemic illnesses. However, most research will require therapy and illness control.

Keywords: Safety, ISO-50001, management, neural networks, remote locality

INTRODUCTION

The danger of pandemics has been a worry for the health care sector throughout history. A contentious issue is the risk that big infectious illnesses may spread across the globe before anybody is aware of it. The apparent predominance of Severe Acute Respiratory Syndrome (SARS) and other kinds of infections in the past have shown how much a pandemic illness may affect nations' health systems [1]. Coronavirus (COVID19) is the last pandemic illness to have a significant effect on the globe. Coronavirus (2019-nCoV) is an infectious illness that was caused by coronavirus 2 (SARS-CoV-2), which started in Wuhan, China on 8 December 2019 [2]. Because a novel coronavirus (nCoV) is an unprecedented new strain in the coronavirus family, the globe confronts significant difficulties to manage this epidemic [3]. In the course of fertile epidemics not only clinical experts tried to develop new treatments and vaccines but also scientists in the field of data science and technology tried to detect the infectious disease and assist it control using information-based methods [4]. Today, large amounts of health information are gathered from many sources through patient care because of the digital health revolution. Therefore, the contemporary medical world is abundant in information but lacks knowledge [5]. Striving to achieve this new pandemic and potential future pandemics has thus become one of scientists' main concerns. In recent decades several important research on pandemics and data mining (DM) methods have been published [6]. These research were carried out with the objective of better understanding, regulating and managing pandemics utilising different techniques of data mining. Because the pandemic COVID-19 is important, a study of the most popular and efficient ways of data extraction may have a significant influence on choosing the most effective strategies of pandemic research. It may therefore assist us to show the unknown nature of the current pandemic and the future potential pandemic. This study aims to gather, summarise and analyse the available papers in order to monitor and analyse research published in terms of pandemics and data mining techniques. Specific research questions (RQ) of this review are: (RQ1) To determine how many studies have been published in recent years and months on the last pandemics and outbreak of COVID-19; (RQ2) To provide overviews and their characteristics; (RQ3) To study the published studies on data mining techniques; (RQ4) Data sources identification (RQ5).

***Corresponding author**
Mohammed Hamid Abood,
Al kunooze University College, Iraq
e-mail: mohammed.h@kunoozu.edu.iq

PHASE OF DATA EXTRACTION

311 publications have been found on scientists' websites in scientific databases (Web of Science, Scopus and PubMed). For screening articles, certain inclusion and exclusion criteria have been defined. All titles and abstracts of the publications were reviewed in the first step to identify suitable research. All the titles and abstracts were checked for the relevant papers by three reviewers (MT, SS and SR). Another examiner (MG) has randomly examined a selection of studies. The Joanna Briggs Institute (JBI) has evaluated the quality analysis of the publications, which offers rigorous criteria for the evaluation of most studies[7]. Since we considered all kinds of research in our evaluation, we used this checklist. Two reviewers made decisions on research eligibility and quality; any discrepancies were settled by debate. Fig. 1 shows the fow of projection articles based on the [8] PRISMA technique.

Phase 3 includes complete text screening. In this step four reviewers carefully reviewed the whole texts of the relevant research (MT, SS, SR, and MG). Through a full-text review, RS took a final judgement if the authors disagreed when selecting eligible papers.

Finally, 50 studies remained qualified. Certain classifications were supposed to categorise and analyse the research presented. Researchers developed the extraction forms to handle the reviewed papers. This classification includes broad information and specific details. The general information includes authors, publishers and dates. Specific information covers the primary goals, DM methods, DM method implementation, health discipline, major outcomes, evaluation findings, sources of information, sample sizes, applicable software and the nation.

Articles included have been evaluated to extract their features on the basis of a predefunction. All the information collected was re-examined to achieve agreement by all the writers. The following reviewer (RS) has assessed and verified the findings. EndNote X9 is utilised for resource management and SPSS v20 does all qualitative analyses.

RESULTS AND DISCUSSIONS

The primary goal of this study was to synthesize the research conducted on data-driven pandemic DM approaches. Consequently, 50 papers from 311 retrieved research have been chosen and evaluated. The findings and finding are described in this section. The data sources utilised were quite varied in the included research. Most studies have been carried out in China. This is due to the fact that most pandemics have started in this nation.

Social media has now become a new source of data [9] and more information may be generated than previous resources within a short period of time. Since accessibility for such data is simpler than other data sources, the most important research focused on the application of text mining methods for information communication. The qualitative study showed that scientists favoured using supervised methods such as regression to build prediction models in order to comprehend unknown pandemics more effectively. All these techniques were pragmatically applied efficiently in many fields of medicine [10]. Classification techniques have also been utilised more than anticipated in research. The easiest way to apply prediction models is for scientists to identify specific biomarkers in unknown illnesses, which enable them to anticipate significant outcomes [11]. The development of forecasting models may thus assist not just medical practitioners but also health policymakers and society.

Since the bulk of research have been carried out in China, these models may be overcome. None of the research suggested that created models be used in actual life. However, majority of the writers were hopeful about predictive model development. The viewpoint of Shamsuddin on the evolution of forecast modelling is consistent with our research [12]. Wyntass et al. performed a comprehensive analysis of COVID-19 prediction models. They found that the models presented were not well documented with a significant possibility of partiality. Results have shown that limiting infectious illness transmission is the primary issue for pandemic sickness. Usually the nature of a new pandemic illness is unknown, and it is one of the scientists who are most concerned to identify the features of a new disease.

This is why the bulk of research focus on disclosing the features of the illness. The fact that scientists should be more diagnosed than other jobs of pandemic illness may explain it. The second major problem with pandemic illnesses is how the disease spreads. Therefore, nearly 10% of research focused on forecasting the prevalence of the illness.

Due to a range of techniques, however, the sample size of the data sets is extremely varied. The findings revealed that most research have utilised different data sources with a restricted number of data sets. The use of big databases may increase the strength of the findings and enhance the predictions of the model, which can in turn assist scientists to better understand this new illness. Researchers are thus advised to utilise big data sets globally for their studies, to obtain better diagnostic and treatment choices. The majority of efforts in terms of illnesses were performed under title COVID-19. Second, the issues linked to pandemics of influenza. The high incidence of these two illnesses is anticipated to ensue. The use and retrieval of vast quantities of data supplied as a data source by electronic systems may enhance data access [13]. As a consequence, data-driven research have become simpler than ever in recent years. The reason that illnesses associated with previous pandemics did not arise in this

*Corresponding author
Mohammed Hamid Abood,
Al kunooze University College, Iraq
e-mail: mohammed.h@kunoozu.edu.iq

search may be owing to epidemics considered by the writers of these publications. We have met several constraints in this research. Currently, a large majority of research on COVID-19 are released everyday. The literature was examined till 16 Oct 2020. Some research may thus be ignored at the time this article is published. Further study is thus required to complete our findings. The suggested study also limits the electronic search procedure to just 3 journal databases, and the other databases were ignored when examining the quality of journal articles which may be addressed in future research. This study provides researchers with a helpful foundation for future work in order to comprehend the broad context of data extraction methods in pandemics and their applications. Further research may include a more broad-based analysis of data mining applications or the development of search methods in bigger datasets.

In order to classify the data into the required clusters, a Feed Forward neural Network (Figure 1) is utilized to develop a smart paradigm. This paradigm may predict occurrence of research activity depending of the attitude of data cluster. Model employing the feed forward neural network for predicting of the same.

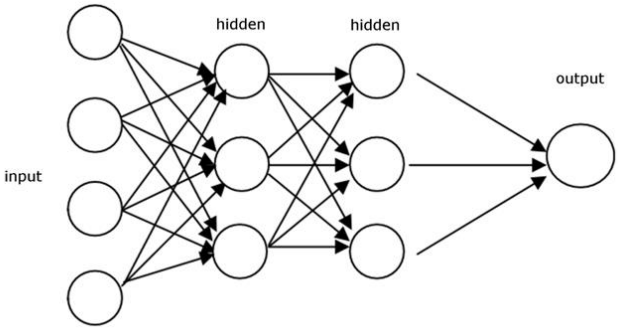

Figure 1: Artificial neural network (ANN) structure

Table 1: Feed Forward model configuration

Settings	Value
Hidden layers count	One
Output layer count	One
Input layers count	One
Algorithm of training	LM
Performance metric in training	Mean Absolute Error
Targeted MAE	$1.009\ e^{-1000}$

Knowing that using the smart technology may require a knowledge of the side effects or the challenges associated with the same. Data mining integration with shopping platform is appeared to be advantageous in many concerns such as marketing, shopping, auditing, etc. From the other hand, once have to think on the cost of using the smart technology over the shopping platforms and which specific technology is fitting the requirement of them.

Data mining role is become necessary in delivering efficient shopping experience for the customers and to ensure high profitability for the service providers. In the preceding sections, a detailed about mining the customer behaviors over the shopping platform as well as details of the event prediction and management system are provided. The implementation of those requirements required innovation of classification, clustering and prediction algorithms. From those available algorithms, are K-means, Random Forest, Neural network, etc. Deep Learning algorithms are also dependable such as Deep Neural Network (DNN) and Long short Term Memory Neural Network (LSTM).

Most of the mentioned algorithms are complex and required a particular hardware availability (high power processors and RAMS on the server side). This might impose the challenge of cost of adaptation and cost of maintaining the same. Another technique which provide simple and straightforward procedure to construct the systems required in e-shopping portals. Such is called as IF-ELSE programming, or in other word, the simple programming codes utilization to construct shopping enhancement logics. This type of codes is much cheaper in cost than the others and possible to be used as well.

The dispute of which technology is more useful for e-shopping platform can be end by addressed by understanding the volume of business. For large volume of business, it is always preferable to deploy highly equipped servers along with most advance technology. However, data mining techniques can also preserve for the small volume business using the called on cloud server. In

*Corresponding author
Mohammed Hamid Abood,
Al kunooze University College, Iraq
e-mail: mohammed.h@kunoozu.edu.iq

recent years, cloud servers can be used to host the shopping portal setup and provide all the facilities of data mining on-cloud on rental bases.

CONCLUSION

This review may assist scientists to facilitate published research on DM methods and ferce pandemics. In this research, we examined the methods used in data mining in global pandemics, although most of those strategies have been created to prevent and forecast the COVID-19 outbreak in the present environment. According to our survey, the main aim of DM applications was the features of the illness. It may also assist politicians and decision-makers to better decide how to manage and avoid severe pandemics in nations.

REFERENCES

[1] Jain V, et al., "Planning for large epidemics and pandemics: challenges from a policy perspective", Curr Opin Infect Dis, 2018.
[2] Cook AH, et al., "Pandemic Disease: A Past and Future Challenge to Governance in the United States", Rev Policy Res, 2008.
[3] Cascella M, et. al., "Features, evaluation, and treatment coronavirus (COVID-19)", StatPearls [Internet], StatPearls Publishing, 2020.
[4] Cucinotta D, et al., "WHO declares COVID-19 a pandemic", Acta bio-medica: Atenei Parmensis. 2020.
[5] Meo SA, et al., "Biological and Epidemiological Trends in the Prevalence and Mortality due to Outbreaks of Novel Coronavirus COVID-19", Journal of King Saud University – Science, 2020.
[6] Zhang S, et al., "Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis", Int J Infect Dis, 2020.
[7] Atif I, et al., "The Role of Digital Technologies that Could Be Applied for Prescreening in the Mining Industry During the COVID-19 Pandemic", Transactions of the Indian National Academy of Engineering, 2020.
[8] Gholamzadeh M, et al., "Suggesting a framework for preparedness against the pandemic outbreak based on medical informatics solutions: a thematic analysis", The International Journal of health planning and management, 2021.
[9] Salzberger B, et al., "Successful containment of COVID-19: the WHO-Report on the COVID-19 outbreak in China", Springer, 2020.
[10] Sun K, et al., "Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study", The Lancet Digital Health, 2020.
[11] Heymann DL, "COVID-19: what is next for public health", The Lancet, 2020.
[12] Keesara S, et al., "Covid-19 and Health Care's Digital Revolution", N Engl J Med, 2020.
[13] Gulyaeva M, et al., "Data mining and model-predicting a global disease reservoir for low-pathogenic Avian Infuenza (A) in the wider pacifc rim using big data sets", Sci Rep, 2020.

*Corresponding author
Mohammed Hamid Abood,
Al kunooze University College, Iraq
e-mail: mohammed.h@kunoozu.edu.iq