



# Insights from the COVID-19 Pandemic: A Survey of Data Mining and Beyond

Imad Afyouni<sup>1</sup> · Ibrahim Hashim<sup>1</sup> · Zaher Aghbari<sup>1</sup> · Tarek Elsaka<sup>2</sup> · Mothanna Almahmoud<sup>3</sup> · Laith Abualigah<sup>4,5,6,7</sup>

Received: 30 January 2024 / Accepted: 1 June 2024 / Published online: 22 June 2024  
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

## Abstract

The global health crisis of COVID-19 has ushered in an era of unprecedented data generation, encompassing the virus's transmission patterns, societal consequences, and governmental responses. Data mining has emerged as a pivotal tool for extracting invaluable insights from this voluminous dataset, offering critical support for informed decision-making. While existing surveys primarily explore methodologies for detecting COVID-19 in medical imagery and official sources, this article comprehensively examines the pandemic through big data mining. We emphasize the significance of social network analysis, shedding light on the pandemic's profound influence on community socio-economic behavior. Additionally, we illuminate advancements in diverse domains, encompassing behavioral impact analysis on social media, contact tracing implications, early disease screening through medical imaging, and insights derived from health-related time-series data analytics. Our study further organizes the literature by categorizing it based on data sources, dataset types, analytical approaches, techniques, and application scenarios. Finally, we delineate prevailing challenges and forthcoming research prospects, charting the course for future investigations.

**Keywords** COVID-19 · Social data mining · Forecasting · Medical imaging · Contact tracing · Time series analysis

## Introduction

Since the recent outbreak of COVID-19, many scientists and big companies have started working on many challenges related to how to fight such an epidemic better, but also how to fight the spread of false information (Zhao et al., 2020; Mavragani, 2020; Cuan-Baltazar et al., 2020). These data are extremely valuable for conducting experiments (Chen et al., 2020; Gozes et al., 2020) for prediction purposes (Lin & Hou, 2020). Particularly in social data mining, some recent work has made a dataset of hundreds of millions of multilingual COVID-19 tweets with location information

---

Extended author information available on the last page of the article

available towards leveraging a collective contribution to extract relevant knowledge on the topic (Qazi et al., 2020). Tracking social media data about the COVID-19 pandemic would help map disease evolution geographically and predict communities under potential risk. It can also help limit the spread of false and misleading information between users, referred to as Infodemics Mavragani (2020); Cuan-Baltazar et al. (2020). Recently, Facebook, Google, LinkedIn, Microsoft, Twitter, and YouTube have committed to removing coronavirus-related infodemics and misleading news due to the severe damage that is caused to human health and society (Cuan-Baltazar et al., 2020).

Social media recently appeared as a great support for understanding the behavior of users and communities. Analyzing crowdsourced data can provide deep knowledge about surroundings, including current social events or unusual happenings. Topics of interest and daily events are heavily discussed over social networks as users share their feedback, multimedia content, and check-ins of their visited places. With the advancement of social and mobile sensing technologies, there is a real opportunity to enrich current detection systems with knowledge extraction tools that leverage tracking and predicting epidemic diseases, such as the COVID-19 outbreak. Moreover, social media data is becoming a rich source of information that can be utilized in detecting and predicting communities under potential risk.

The research focus of this study revolves around the generation and analysis of COVID-19 related data, particularly emphasizing the role of data mining techniques. Our study is driven by the huge available social data shared by people over social networks about topics of interest that can be considered as a triggering point of investigation for the system. For instance, 54 million tweets about COVID-19, collected from Feb.1 until May 1, 2020, were made publicly available recently for exploration (Sharma et al., 2020). However, spreading false information and various harmful content over social media threatens the whole online social ecosystem. The ongoing COVID-19 pandemic presents no exception by laying the ground for an army of malicious users to spread low-credibility and unverified news.

Besides, the COVID-19 global pandemic has resulted in a massive global disruption in the healthcare sector and the economy, education, environment, and social life, to name a few (Leung et al., 2022). Besides the battle on the medical front line, governments, industries, and the research community have extensively explored the deployment of information and communication technologies to track and contain the global outbreak. Mobile Contact Tracing Applications (MCTA) were developed as part of these efforts, which take advantage of the vibrant ecosystem of mobile sensing (e.g., location, proximity) to leverage identifying and tracking pedestrians that may be contagious or under potential infection threat. Although contact tracing apps have raised many important privacy concerns, such technologies will put in place a massive global surveillance infrastructure that may survive even after the partial containment of the disease due to the worldwide distribution of the vaccine.

Figure 1 illustrates the steps of COVID-19 data analysis concerning social media, medical imaging, time series data, and contact tracing. Various data sources are available to extract insights and generate knowledge layers from raw data, such as geo-tagged tweets, time series of numerical daily data, X-ray images and topics and reviews from online platforms, and data from contact tracing apps. After pre-processing such

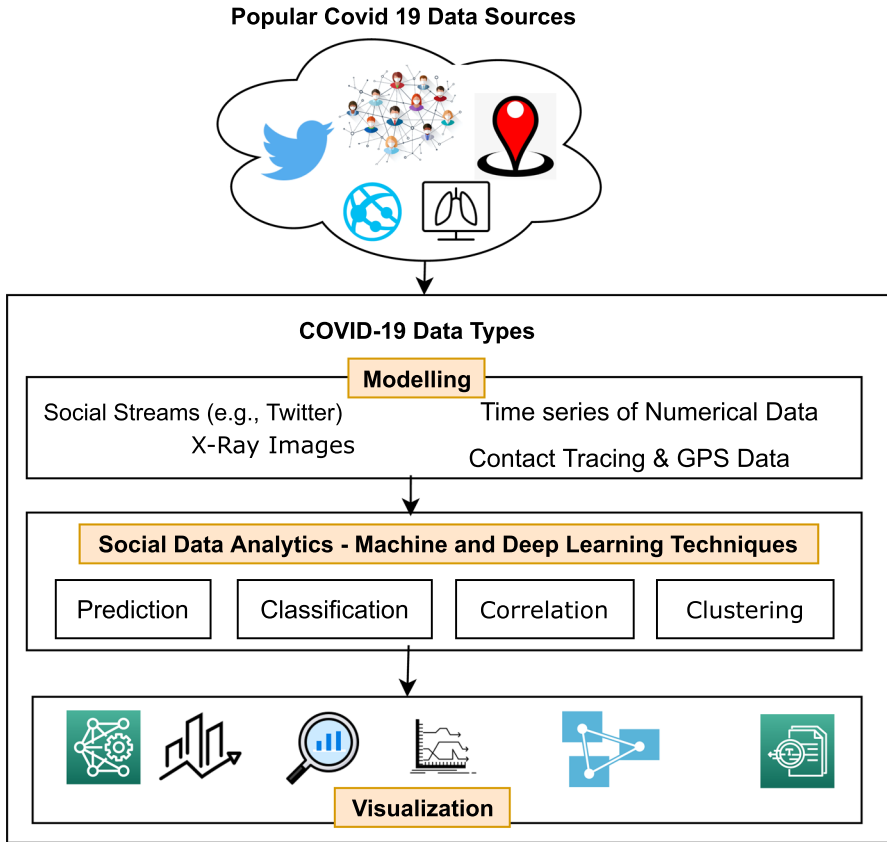


Fig. 1 Overview of COVID-19 data analysis

data, they can be utilized for various types of analysis, including prediction, classification, correlation, and clustering. Finally, modeling and visualization tools summarize the data for decision-making by the concerned authorities.

The main contributions of this survey are: 1) introduce a taxonomy of related work on COVID-19 data mining techniques and analytics from each respective domain; 2) discuss the main data sources and datasets that have been recently produced to facilitate mining tasks, with a focus on social data; 3) present data modeling foundations, and then overview techniques in social data mining, medical imaging, contact tracing, and time-series data; 4) summarize analytical perspectives and implications in each research track, and the impact of COVID-19 on the socio-economic behavior; and finally discuss the overall challenges and great opportunities in these domains.

**Related Work**

This paper is motivated by the growing number of studies conducted on the COVID-19 pandemic. The most relevant work may appear in Chiroma et al. (2020) and Shinde et al. (2020). These papers offer a comprehensive discussion on the analysis of COVID-

19 data extracted from public official websites or medical images. However, the main focus of this paper is COVID-19 social media data analysis. Our paper surveyed the existing literature on COVID-19 social analytics from various aspects and discussed their challenges and opportunities. Other related work can be categorized as contact tracing, prediction, and economic impact.

(i) **Bibliometric analysis:** Chiroma et al. (2020) provided a survey on early assessment using bibliometric analysis based on a machine learning approach to limit the spread of COVID-19. The authors have collected the dataset from academic databases and bibliometric techniques for analysis. Moreover, a new perspective is proposed to overcome some of the challenges highlighted. Their results indicate that using machine learning-based COVID-19 diagnostic tools would require considerable attention. Lazarus et al. (2021) surveys potential acceptance of a COVID-19 vaccine. The main goal is to collect data from various nations in order to determine the possible acceptance rate and factors impacting COVID-19 vaccine acceptability. The study concluded that 48% of the people surveyed would accept the vaccine if recommended by their employers or the government.

(ii) **Medical image:** A survey on deep learning and medical image processing for COVID-19 was presented by Bhattacharya et al. (2021). The authors summarize the recent research work related to deep learning and its application to healthcare. Moreover, three use cases concerning China, Korea, and Canada have been studied to confirm the uses of deep learning applications for COVID-19 medical image processing. Also, the study highlighted some of the challenges related to deep learning implementation for COVID-19.

The authors in Ulhaq et al. (2020) surveyed the proposed computer vision techniques for COVID-19 control. They discussed the recent methods: computed tomography (CT) scans, X-ray Imagery, and Prevention and Control. The authors also identified some of the future research directions concerning the COVID-19 pandemic. Besides, Ahmed et al. (2020) has provided a comprehensive COVID-19 Contact Tracing Apps survey. The authors also present an overview of many proposed tracing app examples and discuss users' concerns regarding their usage.

(iii) **Other related work:** Mahalle et al. (2020) provide a survey on forecasting models for COVID-19. The authors classify the forecasting techniques into two types, which are (i) stochastic theory mathematical models and (ii) machine learning techniques. Moreover, the authors highlighted some of the challenges and recommendations of the forecasting techniques. Alamoodi et al. (2020) studied and examined papers from the last ten years regarding the prevalence of various forms of infectious diseases, such as viruses, epidemics, pandemics, or outbreaks, understand the use of sentiment analysis, and collected the most significant literature findings. From January 1, 2010, to June 30, 2020, they systematically searched papers on similar topics in major databases. They organized the papers into a taxonomy to classify the current literature's corresponding current viewpoints in four categories: lexicon-based models, machine learning-based models, hybrid-based models, and individuals. They divided the publications they found into three categories: disease mitigation, data analysis, and issues researchers face with data, social media platforms, and community. They discovered some interesting patterns in the literature and categorized the articles accordingly.

The previous surveys focus mainly on COVID-19-related medical Bibliometric, forecasting, and image data analytics or data extracted from official platforms. However, our study aims to offer a comprehensive survey on COVID-19 from social data mining perspectives. More specifically, the study covers the literature on the social impact of COVID-19, including data sources, social media analytics, contact tracing, prediction, and its impact on the economy. Lastly, we highlight some challenges and opportunities to address as future research directions. Taxonomy of COVID-19 Data Analysis is given in Fig. 2.

### Taxonomy and Search Methodology

The taxonomy devised for comprehensively analyzing existing studies related to COVID-19 data analysis from a social perspective encompasses five key categories: data sources, type of datasets, type of analytics, techniques, and type of use cases. Each category provides a framework for understanding the various aspects of COVID-19 data analysis undertaken in the past two years. Data sources range from official government sources to social media platforms, while datasets include epidemiological, demographic, geographic, healthcare system, mobility, and socio-economic data. Analytical methods span descriptive, predictive, and prescriptive analytics, as well as network analysis, text mining, and machine learning techniques. Specific methodologies such as sentiment analysis, contact tracing, social media analysis, prediction are discussed, alongside diverse use cases such as public health interventions, resource allocation, risk assessment, economic impact analysis, and mental health assessments. This taxonomy facilitates a structured approach to evaluating the multifaceted nature of COVID-19 data analysis and informs future research directions aimed at addressing societal challenges posed by the pandemic.

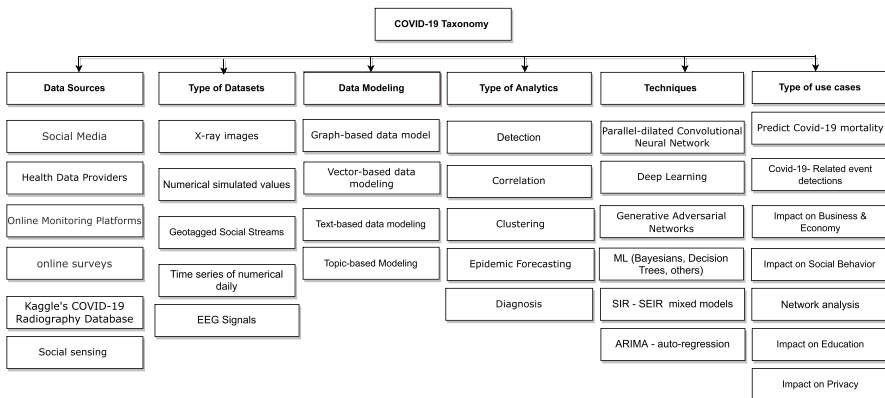


Fig. 2 Taxonomy of COVID-19 data analysis

## Data Collection

Scopus and Google Scholar are widespread databases that cover the comprehensive scientific data and literature (Boyle & Sherman, 2006). Hence, the references covered in this study were obtained from Scopus and Google Scholar databases. We gathered bibliographical information, citations, abstracts and keywords, and funding details relevant to COVID-19 research from 2020 to April 2022. We used keywords separated by Boolean operators AND, OR, and NOT such as ( TITLE-ABS-KEY ( "Social Media" ) AND TITLE-ABS-KEY ( analytics ) OR TITLE-ABS-KEY ( analysis ) OR TITLE-ABS-KEY ( data AND mining ) AND TITLE-ABS-KEY ( covid-19 ) OR TITLE-ABS-KEY ( coronavirus ) ) AND PUBYEAR > 2019 to retrieve specific literature. Moreover, we omitted the term "sars-cov-2" from the search because it returned 3,500 documents; some of the documents were beyond the scope of this study. Journal articles, conference papers, reviews, book chapters, Notes, Editorials, and short surveys are retrieved from the database. The final dataset presented in this study was filtered from 3,500 bibliographic entries published between January 2020 and April 2022.

## Data Sources

The primary data sources about COVID-19 were social media, Internet Search Engines, health data providers, online monitoring platforms, and other data providers, such as government websites and international organizations. Social media platforms such as Twitter and Facebook provide valuable data on public sentiment, behaviors, and the spread of information. In the subsections, we discuss the different types of data sources and their COVID-19-related contributions. Table 1 classifies the papers based on the data sources.

## Social Media

Social media is the main data source that can reflect users' behavior and opinions about topics of interest. The work in Koh and Liew (2020) studied the effect of COVID-19 and social distancing on loneliness and other mental health issues using Twitter data. In Mutlu et al. (2020), Twitter data is used to extract the users' opinions on using two medications (i.e., hydroxychloroquine and chloroquine) to cure COVID-19. Investigated anomalous social movement during COVID-19 (Devi & Nayyar, 2021) through sentiment analytics of geo-tagged tweets. The work in Al-Rawi and Shukla (2020) focused on the impact of bots activities, which are programmed accounts that tweet or retweet mentioning the hashtags #COVID-19 or #COVID-19. Alsudias and Rayson (2020) analyzed 1M Arabic tweets to detect rumors and predict the source of the tweets. The authors of Ordun et al. (2020) evaluated the effect of using certain features, keywords, and some unique topics, as well as how fast the COVID-19-related information is being tweeted or retweeted. In Chakraborty et al. (2020), the authors discuss that official COVID-19 platforms such as WHO were not successful in precise

**Table 1** Paper classification based on their data sources

Data source	Articles	Remarks
Social media	Koh and Liew (2020), Mutlu et al. (2020), Devi and Nayyar (2021), Al-Rawi and Shukla (2020), Alsudias and Rayson (2020), Ordun et al. (2020), Chakraborty et al. (2020), Zheng et al. (2020), Samuel et al. (2020), Abd-Alrazaq et al. (2020), Gencoglu et al. (2020), Haouari et al. (2021), Qazi et al. (2020), Alqurashi et al. (2020), Lamsal (2020), Kabir and Madria (2021), Dimitrov et al. (2020)	Most of the work is collecting datasets and then doing some analysis on these datasets.
Internet search engines	Li et al. (2020), Lai et al. (2020), Li et al. (2020)	These works Used user searches on Google, Baidu and Weibo to predict the daily cases and spread analysis.
Health data providers	Liu et al. (2020), Hos-sain et al. (2020), Wang et al. (2020), Amram et al. (2020), Chao et al. (2021), Alzahrani et al. (2020)	Some papers studied the impact of quarantine on the increase of Coronavirus cases and the spread of the disease. Other papers used hospital data to predict daily cases and the spread of COVID-19.
Online monitoring platforms	Russo et al. (2020), Bhat-tacharjee (2020), Nadim et al. (2021), Rocha Filho et al. (2020), Traini et al. (2020), Anastassopoulou et al. (2020), Giordano et al. (2020), Bayham et al. (2020), Beare et al. (2020), Siwiak et al. (2020)	These papers used data published by WHO, John Hopkins University, and government official websites to analyze and predict the COVID19-related daily cases, deaths and spread rate.
Other types of data providers	Rovetta and Bhagavathula (2020), Lai et al. (2020), Zhu et al. (2020), Kiamari et al. (2020)	Some papers used news media to predict the increase of COVID-19 cases. Other papers used mobile data, e.g. travel data, to study the impact of the travel restriction on the spread of COVID-19.

and informative information to guide the public and thus reduce the spread of the disease. The work in Zheng et al. (2020) used topic modeling to classify tweets into discussion topics about COVID-19. Samuel et al. (2020) identified public sentiment, e.g., fear sentiment over time, developed due to Coronavirus. However, Abd-Alrazaq et al. (2020) classified the tweets into different COVID-19 pandemic topics.

Gencoglu et al. (2020) proposed a machine learning-based classification to categorize tweets based on their language-agnostic representations. Section 5 presents a detailed analysis of these techniques and classifies them based on their detection purpose, language covered, and datasets used.

### Internet - Search Engines

studied the propagation and spread analysis of COVID-19 in China Li et al. (2020) and Lai et al. (2020) using data collected from Google and Baidu search engines. In Li et al. (2020), the authors presented an analysis using Google Trends, Baidu Index, and Weibo Index to conclude that reported daily cases are highly correlated with social media posts as well as Internet searches.

### Health Data Providers

The work in Liu et al. (2020) discussed the impact of COVID-19 control measures on the spread of the disease. Another studied parameter related to the spread of COVID-19 is the movement restrictions. They used a dataset from the US Centers for Disease Control (CDC). While Hossain et al. (2020) used data collected from the Chinese Center for Disease Control and Prevention to study the impact of quarantine on the spread of COVID-19. The authors in Wu et al. (2021) studied China's responses to the COVID-19 emergency from a policy-making perspective since the beginning of the outbreak by adopting a joint unit for epidemic prevention and control mechanisms. Moreover, Wang et al. (2020) used data from the Centers for Disease Control, China, to predict the spread rate of COVID-19. A study to identify at-risk individuals using data collected from the International Classification of Disease was conducted by Amram et al. (2020). On the other hand, Chao et al. (2021) discussed the use of imaging and non-imaging data collected from hospitals in Iran and Europe of COVID-19 patients to predict patient need for ICU admission. While Alzahrani et al. (2020) presented a method to predict daily cases in KSA by analyzing the official data from the Ministry of Health.

### Online Monitoring Platforms

Researchers have used official statistics to estimate and predict diseases with negligible biases and small computational requirements. The work in Russo et al. (2020) analyzed COVID-19 databases accessed from WHO. Moreover, Bhattacharjee (2020) studied the impact of local environmental factors, such as humidity and temperature, on the spread of COVID-19. Also, based on WHO data, Nadim et al. (2021) and Rocha Filho et al. (2020) discussed the effect of quarantine on the spread of the Coronavirus. Using publicly available data from the Hubei province in China, the work in Anastasopoulou et al. (2020) predicted the number of new cases in China. On the other hand, Traini et al. (2020) and Giordano et al. (2020) collected data from the Italian National Data to study the coronavirus spread and death rates. Used data on the GitHub page



(Bayham & Fenichel, 2020) to show how social distancing affected the number of death cases. Infection and recovery rates were studied from data downloaded from GitHub, uploaded by John Hopkins University (Beare & Toda, 2020; Siwiak et al., 2020).

## Other Types of Data

News media TV/video, newspapers, and radio were explored by some works to extract knowledge related to COVID-19. Additionally, the authors in Rovetta and Bhagavathula (2020) used mobile data of people in Italy and online searches to study the impact of some discussion topics on the spread to the Coronavirus. In Lai et al. (2020), the authors used the mobile data of domestic and international travel to study the impact of travel on the increase and spread of Coronavirus cases. However, the work in Zhu et al. (2020) used mobile data to predict the number of death cases by modeling the decay rate of the spatial mobile data. To estimate risks upon lifting lockdown, the authors of Kiamari et al. (2020) developed a Hawkes process-based technique that uses cell-phone-based mobility data to compute spatiotemporal risk scores assigned.

## Techniques

We discuss in this section the main techniques developed recently to extract insights from diverse types of data related to COVID-19. Our main focus will be on covering the social data mining techniques. Still, we will also highlight the recent contributions in different related fields, such as medical imaging techniques for early screening and classification of COVID-19 cases and contact tracing mobile applications that were proposed in many countries by officials to track and help prevent the spread of the disease. This section first presents some foundations on mathematical modeling and methods for representing data related to COVID-19, then discusses the social data mining techniques in different languages. In addition, we present the techniques used in medical imaging and contact tracing and finally show assessments of these techniques based on well-established evaluation metrics whenever available.

## Mathematical Data Modeling

Researchers use mathematics and computer tools to model the patterns of the COVID-19 pandemic. These models aim to try to understand the patterns, predict future outbreaks, and track the evolution of COVID-19. In Tang and Wang, (2020), the authors modeled the decrease of the daily growth as an exponential decay function. The work in Kucharski et al. (2020) used mathematical models to assess the human-to-human transmission of COVID-19 in different areas in China. In Oehmke et al. (2021), the authors used a dynamic panel data model estimated using the generalized method of moments approach to provide surveillance metrics for COVID-19. These metrics provide estimates for speed, acceleration, weekly shifts, etc., to support decision-making to alleviate risks. COVID-19 cases and mortality data, along with a deterministic SEIR compartmental framework, are used by COVID et al. (2020)

**Table 2** Social data mining during the COVID-19 pandemic

Reference	Purpose of study	Detection technique	Type of analytics	Datasets	Emerging technologies	Evaluation metrics
Park (2022)	Impact of COVID19 on crisis management	Topic modeling from Twitter feeds	Big social data and its effects on crisis management	Arabic news and Twitter/YouTube feeds	Unsupervised ML	55
Mutlu et al. (2020)	Mining opinions on drug usage for prevention of COVID19	Stance annotation/sentiment analysis and manual Polarization among audiences	Use of 'chloroquine' and 'hydroxychloroquine' for COVID19 treatment	Twitter data (14,374 original tweets) for April 2020	Supervised ML, Classification methods: MLP, LR, SVM, MNB, SGD, GB, and CNN	Accuracy using tf-idf
Devi and Nayyar, (2021)	Detection of social sentiment	Sentiment analysis and classification	negative, positive neutral tweets	6,468,526 tweets	ML	55
Qazi et al. (2020)	Large-scale Twitter dataset	Gazetteer-based approach to extract toponyms from user location and tweet content	Geographic coverage of data, top countries and tweet distribution	524 million multilingual tweets	Location Inference from Text	55
Ordun et al. (2020)	Trending topics, causal events, impact on social behavior	Topic modeling and clustering using LDA and UMAP	Top Topics, trends over time, average frequency of keywords, time to retweet	5,506,223 tweets	55	
Lamsal (2020)	COVID19Tweets dataset design	sentiment scores using LSTM	Daily distribution of tweets, network graph and sentiment trends	310 million English tweets	NLP, LSTM	55

Table 2 continued

Reference	Purpose of study	Detection technique	Type of analytics	Datasets	Emerging technologies	Evaluation metrics
Sharma et al. (2020)	Analyzing misinformation in Twitter	Used fact-checking sources, sentiment analysis and classification	Identified unreliable and misleading content, spreading patterns of misinformation tweets, distribution of engagements	54.32M tweets	Real-time tracking of misinformation	F1 and accuracy metrics
Kabir and Madria (2021)	EMOCOV dashboard for topic trends and human emotions in USA	Sentiment analysis and data analytics	Frequent topics, sentiment distribution, user movement correlation, and polarity over time	20M tweets	NLP, Sentiment Analysis	55
Chakraborty et al. (2020)	Sentiment analysis and impact of popularity on fact spreading in social media	Deep Learning Classifiers and Gaussian membership	Topic trends, sentiment and misleading information analysis	226,668 tweets	Deep learning and word embeddings	55
Gencoglu (2020)	Twitter discourse classification	text classification, Bayesian optimization, multilingual sentence embeddings using LaBSE	UMAP visualization of sentence embeddings, Normalized confusion matrix of predictions	26 million tweets	natural language processing, deep learning	F1 and accuracy metrics

Table 2 continued

Reference	Purpose of study	Detection technique	Type of analytics	Datasets	Emerging technologies	Evaluation metrics
Abd-Alrazaq et al. (2020)	sentiment analysis and topic surveillance	n-grams, topic modeling using LDA, rule-based classification	Source of the virus, impact on people and countries, methods for decreasing COVID-19 spread	2.8 million tweets	unsupervised machine learning	highest/lowest means
Hou et al. (2021)	Public attention and users' sentiments on popular topics	Weibo texts, topic modeling using LDA, probability distribution, Spearman correlation	public attention level trend	1.53K Weibo texts	unsupervised machine learning	Correlation coefficient

to model trajectories of severe acute respiratory syndrome SARS-CoV-2 infections. Using the model, the authors assessed social distancing and mask use levels on the virus's spread.

Mathematical modeling Oliveira et al. (2021) to study the dynamics of COVID-19 in Bahia, a state in northeastern Brazil, considering the influences of asymptomatic/non-detected cases, hospitalizations, and mortality. The model explored hospitalization needs in a low-resource state during the COVID-19 pandemic. Mathematical modeling of COVID-19 data can lead to radical shifts in government decision-making Ferguson et al. (2020). For example, governments implementing 'herd immunity' strategies had to change their strategies after mathematical models predicted enormous death rates before reaching this objective.

Data-driven compartmentalized (susceptible-infected-recovered) modeling provides insights into the spread of COVID-19. Recent models that experimented with an increased number of compartments studied the impact of social distancing and quarantine on the spread of the virus and other statistics such as number of daily cases, deaths, etc. Leung et al. (2020) Giordano et al. (2020). Investigations into the need to hospitalize patients of COVID-19 in different scenarios were presented using mathematical modeling of public-related data by Moghadas et al. (2020) Castro et al. (2020).

## Mining COVID-19-Related Insights from Social Media

Many research studies have been presented since the beginning of the outbreak on how to make use of the huge amount of social media streams to extract insights and get a better understanding of the evolution of the disease, the spread of misleading information, and the users' behavior and belief concerning related topics, such as the lock-downs and vaccines, towards overcoming the COVID-19 pandemic Yao et al. (2021). The following will present the different methods used in social data mining, which mainly covers analytics and sentiments on the most frequent topics, the evolution and tracking of the disease over time and space, and the different visualization mechanisms adopted. Tables 2 and 3 illustrate the main contributions in social data mining for COVID-19.

### Large-Scale Datasets

Several developments have been presented on crawling large-scale social datasets to discover and track the evolution of the pandemic and to deeply model user behavior concerning the most trending topics. The design and analysis of a large-scale COVID-19 tweets dataset was introduced in Lamsal (2020). The Twitter dataset has more than 310 million COVID-19-specific English language tweets and their sentiment scores (Lamsal, 2020). They also presented the GeoCOV19Tweets Dataset (Lamsal, 2020), the dataset's geo-tagged version. They analyzed the tweets in both datasets based on trending unigrams and bigrams with scores. Different algorithms for filtering geo-tagged tweets, hydrating tweet data using Twarc, and extracting region-based tweets were presented. They released these datasets publicly.

**Table 3** Social data mining during the COVID-19 pandemic (Cont..)

Reference	Purpose of Study	Detection technique	Type of Analytics	Datasets	Emerging technologies	Evaluation metrics
Haouari et al. (2021)	Arabic COVID-19 Twitter dataset	Collection of source tweets and propagation networks	top general key-words, hashtags, trending topics, distribution of geo-located tweets	748K tweets	51 popular	51
Al-Rawi and Shukla (2020)	Activities of automated social media accounts or bots	Botometer algorithm, topic modeling, Gephi for social network analysis	categories and subcategories of bots' memes, frequent topics and emojis	50M tweets, 185K messages posted by 127 bots	Health Chatbots, Social Network Analysis	55
Shaar et al. (2021)	Crowdsourcing information for infodemic detection	SVM classifier, FastText, BERT for word embeddings	Arabic and English tweet annotations	504 English and 218 Arabic tweets	Online crowd-sourcing annotations	55
Alsudias and Rayson (2020)	k-means clustering, ML classifiers with manual labels on false information	ML and clustering algorithms	identifying main topics, detecting rumours, and predicting tweet source	1M tweets	ML, Arabic NLP	Accuracy, F1, Recall, Precision
Ayoub et al. (2021)	Explainable NLP model to combat misinformation during COVID-19	DistiBERT embeddings and SHAP	Better explainability in detecting fake news	984 verified claims about COVID-19	Word Embedding and Explainable NLP	51
Koh and Liew (2020)	Impact of COVID-19 on loneliness	Topic modeling, hierarchical clustering	identifying key themes related loneliness, community and social distancing impact	4492 Tweets	Natural Language Processing, Machine Learning	Prevalence of themes over time

Table 3 continued

Reference	Purpose of Study	Detection technique	Type of Analytics	Datasets	Emerging technologies	Evaluation metrics
Dimitrov et al. (2020)	Extraction and publishing of TweetsCOV19 semantic datasets		Extension of TweetsKB corpus, semantic annotation and metadata extraction	270 million RDF triples describing metadata for about 8.1 million tweets	SPARQL, Sentiment Analysis, Semantic Annotation	55
Zheng et al. (2020)	Studying temporal differences in topics	LDA topic modeling, Gibbs sampling	evolution of discussions over time, monitor the spread of disease	18.8 million tweets	R language, Topic modeling and tracking	55
Samuel et al. (2020)	Fear-sentiment over time in US	textual analytics and visualization, supervised ML classifiers	Sentiment map, exploratory topic analysis	900K tweets	ML classification and sentiment analysis	55
Hamzah et al. (2020)	CoronaTracker dashboard	predictive modeling, SEIR models	forecast COVID-19 cases, deaths, and recoveries, sentiments on news	data from JHU	Statistical Analysis	55

Dimitrov et al. (2020) developed a knowledge base of semantically annotated tweets on the COVID-19 pandemic, called Tweets-COV19, which contains more than 8 million tweets, October 2019 - April 2020. TweetsCOV19 is a subset of TweetsKB, a public RDF corpus of anonymized data for a large collection of annotated tweets. GeoCoV19 is a dataset of hundreds of millions of multilingual geo-tagged tweets on COVID-19-related topics (Qazi et al., 2020). Data crawling was performed on a period from February 1 to May 1, 2020, with more than 524 million multi-lingual tweets (62 different languages) with around 43 million Twitter users. The geo-location information is essential for many tasks, including disease tracking and surveillance. However, Twitter data has by default, a very small percentage of geo-tagged tweets (generally, between 2% to 5%). Therefore, a gazetteer-based approach was employed, which takes advantage of tweet content and user location to detect toponyms and derive its geo-location based on the Nominatim API from Open Street Maps at different spatial scales. The GeoCoV19 dataset leverages the development of AI-based analytics to predict disease outbreaks and trends and to learn about knowledge gaps and the impact of the global pandemic on the socio-economic life of users, among others.

has collected a large Arabic Twitter dataset on COVID-19 (Alqurashi et al., 2020). They collected tweets in Arabic from January 1, 2020, to April 30, 2020. They used specific keywords and hashtags to collect tweets. They provided preliminary statistics on the dataset. The dataset could help researchers and policymakers study different societal issues related to the pandemic, such as behavioral change, information sharing, misinformation analysis, and spreading rumors. Created a dataset called ArCOV-19 Hamzah et al. (2020), an Arabic COVID-19 Twitter dataset of 2.7M tweets spanning one year, covering the period from January 27, 2020 to March 31, 2020. It includes around 748k popular tweets alongside their propagation over the social network. They showed that ArCOV-19 captures the discussions associated with reported cases since the beginning of the outbreak in the Arab world. Aggregating and analyzing large datasets of diverse data helps in tracking the spread of the virus, identifying hotspots, assessing the effectiveness of interventions, and forecasting future trends. Moreover, large-scale datasets enable the development of predictive models for disease transmission, severity, and outcomes, as well as the evaluation of vaccine efficacy and safety. However, challenges such as data privacy, quality assurance, and interoperability need to be addressed to maximize the utility of large-scale COVID-19 datasets while safeguarding individual rights and ensuring data integrity (Bentotahewa et al., 2021).

### Topic Detection and Monitoring

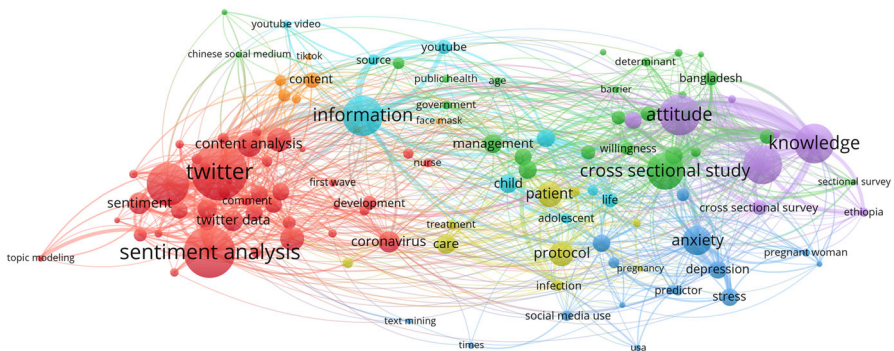
Most of the recent research on social data mining covered some exploratory analysis of COVID-19 streams using topic modeling and detection methods. Most works have collected data from Twitter social networks during the outbreak starting from January or February 2020. For instance, the authors in Ordun et al. (2020) investigated research questions to discover high-level trends and events that can be inferred from COVID-19 tweets. Using UMAP analysis, they inferred local clusters of topics representing personal protective equipment (PPE), healthcare workers, and government concerns. Using document embedding techniques like UMAP allowed for a better understanding



of distinct topics extracted with the LDA method. Topic detection and monitoring are being used by some organizations to understand public sentiment and track discussions related to the COVID-19 pandemic. Such process requires collecting data from various sources such as social media platforms. The main goal is to extract insights to enable decision-making and proactive engagement with healthcare entities. Real-time monitoring and alerting mechanisms ensure timely response to evolving discussions, while regular reporting and actionable recommendations (Organization, W. H., et al., 2021). Wahid et al. (2023) have developed COVICT system demonstrating its potential for early detection, monitoring, and contact tracing. Leveraging real-time symptom data and semi-automated contact tracing can significantly aid in controlling the spread and identifying high-risk areas for targeted interventions. The potential for smart lockdowns and informed policy-making through this IoT architecture shows promise in the ongoing battle against the pandemic.

Research on topic modeling, extraction, and sentiment analysis is increasingly widening its focus as most of the datasets collected from the internet were from popular social media such as Twitter and Facebook. Figure 3 shows the result based on the most frequent terms collected in related papers. Various terms can be seen, such as content analysis, text mining, topic modeling, depression, etc. The minimum number of keyword occurrences used in this study is 5. For each keyword, the total strength of the co-occurrence links with other keywords is calculated. The keywords with the highest total link are selected. The initial term map covering 2020, 2021, and until March 2022 consists of 18 terms in 5 clusters.

An analysis of retweet speed shows that the median retweeting time was approximately 50 minutes faster than repostings from Chinese social media about H7N9 in March 2013. The size of the corpus is 5,506,223 tweets, about 77% of 23,820,322 tweets. From 1 January 2020 until 30 April 2020, Agarwal et al. (2020) developed a framework to classify important tweets relating to the COVID-9 pandemic and have investigated subject modeling to identify the issues and topics most discussed in their data collection. To deal with developments during the pandemic, the authors studied the temporal shifts in the subjects and discovered that eight subjects were enough to classify the themes. These subjects show a pattern tracking over time. Over the years, the dominant themes differ and correlate with the COVID-19 cases.



**Fig. 3** Term map of COVID-19 social media analysis

On the other hand, Kabir and Madria (2021) developed EMOCOV, which uses a collected Twitter dataset to visualize extracted topics and to represent human emotions during the global pandemic. Their dashboard presents various data analytics in the USA over a specified period of time to show changes in topic trends, and human emotions, and subjectivity of user feedback. Abd-Alrazaq et al. (2020) presented an inveillance study on collected data of 2.8 million tweets from 160,829 unique users between February 2, 2020 and March 15, 2020. The tweets were analyzed using word frequencies of single (unigrams) and double words (bigrams). Latent Dirichlet allocation for topic modeling was employed to identify the main topics. Sentiment analysis and interaction rate of topics were performed by extracting the mean number of retweets, likes, and followers for each topic and calculating the interaction rate per topic.

CoronaTracker is an online platform that provides the latest news development, as well as statistics and analysis on COVID-19 Hamzah et al. (2020). They visualized real-time data queries, and then the queried data is used for Susceptible-Exposed-Infectious-Recovered (SEIR) predictive modeling. Their model predicts COVID-19 cases, deaths, and recoveries. It also helps to interpret patterns of public sentiment on disseminating related health information and assess the political and economic influence of the spread of the virus. Among other applications, the authors in Guntuku et al. (2020) also studied the impact of COVID-19-related news on mental health and symptom mentions of users from Twitter data. Moreover, no current work proposes a hybrid social and physical sensing approach to address some of these challenges. Gozes et al. (2020) developed an analysis tool to classify and quantify computer Tomography (CT) images of COVID-19 potential cases using deep learning. Several datasets from disease-infected areas in China were used for the training. Retrospective experiments were conducted to assess system performance in identifying thoracic CT features of COVID-19 potential cases. Zheng et al. (2020) used topic modeling to reveal insights from Twitter users' feedback about the disease. They focused on the temporal analysis of related topics throughout the pandemic. Hou et al. (2021) analysed Weibo texts (from Dec. 2019 to May 2021) to infer the public attention and users' sentiments on 41 popular topics related to COVID-19. Similarly, Zhang et al. (2021) aimed at identifying Twitter groups based on their concerns, sentiments, emotions, and disparities. Text mining from social media to infer policies for healthy and safe airports was also investigated in Park et al. (2021) to enrich the user experience in urban infrastructures.

Sentiment analysis on topics related to COVID-19 has been also studied in other research works. The authors in Chakraborty et al. (2020) analyzed a dataset containing 226,668 tweets collected from December 2019 to May 2020, which contrastingly shows that netizens had a maximum number of positive and neutral tweets tweeted. They demonstrated that though people have tweeted mostly positive regarding COVID-19, netizens were busy engrossed in re-tweeting the negative tweets and that no useful words could be found in word cloud or computations using word frequency in tweets. They validated their proposed model using deep learning classifiers and Bag-of-Words and Doc2Vec models, with admissible accuracy up to 81%. They proposed the implementation of a Gaussian membership function-based fuzzy rule base to identify sentiments from tweets correctly. Sentiment insights on coronavirus-specific

tweets were also studied in Samuel et al. (2020). They demonstrated insights on the progress of fear sentiment over time as the pandemic approached peak levels in the USA, using exploratory and descriptive textual analytics and visualization tools. Their approach discovers early-stage insights using two essential textual classification methods and assesses their ability to classify Corona-related tweets. They observed a high accuracy for classifying short tweets using the Naïve Bayes method. In contrast, the logistic regression classification method yielded a reasonable accuracy with a relatively weaker performance for longer tweets. Nemes and Kiss (2021) also investigated users' emotional polarity from Twitter on COVID-19-related topics using recurrent neural networks and sentiment analysis. Rapid emotional changes and fluctuations were manifested with different classes of emotions and with a good overall classification performance.

### Analyzing Fake News and Misinformation

Analysis and discovery of fake news and misleading information has gained great interest on social media due to the huge incoming unverified streams spread over social networks. This issue is more crucial when discussing health-related topics, especially during the COVID-19 pandemic, where the amount of misinformation shared is colossal (Ayoub et al., 2021). Researchers recently have examined the activities of automated social media accounts or bots and the spread of false news on the pandemic (Nakov & Da San Martino, 2021). The authors in Al-Rawi and Shukla (2020) investigated the activities of social bots by adopting an integrated approach comprising data acquisition, classification/prediction, text mining, and network analysis. They collected data by using tweets or retweets referencing standard terms, e.g., #COVID19, over a period of over two months from February until April 2020. The total sample used was over 50,811,299 tweets from 11,706,754 unique users. The final sample was extracted from more than 185,000 messages posted by 127 bots. They showed the main classes and subclasses of bots' memes. They found that financial incentives drive most bots and try to increase awareness of COVID-19 risks by citing official media and health sources. In contrast, other kinds of bots actively support the survivalist movement by emphasizing the need to prepare for the pandemic and learn survival skills.

Based on this claim, Apuke and Omar (2021) proposed a model for fake-news distribution predictors amongst social media users with Nigeria as a case study. The authors describe the result of a Nigerian sample regarding the dissemination of fake news related to COVID-19. Data was analyzed with Partial Least Squares metrics to find the impact of different parameters on disseminating fake news. An explainable NLP model to detect misinformation from social media was proposed in Ayoub et al. (2021) by using a variant of BERT embedding, DistilBERT, and SHAP (Shapley Additive exPlanations) for better explainability. A dataset of 984 claims about COVID-19 was collected and verified with fact-checking sources and was tested on the COVID-19 dataset. The results show high accuracy in detecting misinformation while figuring out the source of fake news. Analyzing misinformation from COVID-19-related tweets has been also investigated in Sharma et al. (2020). Streaming data from Twitter was collected from March 1, 2020, to June 2020, with 8.1M tweets from 182 countries. They identified unreliable and misleading content based on fact-checking sources and

studied the narratives endorsed in misleading tweets and their distribution of engagement. Misinformation is identified by evaluating the retweet trees of a given post. A statistical dataset of source tweets with labels on misinformation cascades was used, and then a classifier was developed in Sharma et al. (2019) with a character-level embedding to determine suspicious cascades. The dashboard presented analysis and a daily updated list of identified misinformation claims during the pandemic. They provide examples of the spreading patterns of potentially misleading tweets. Various use cases have emerged in analysing fake news and misinformation during COVID-19. One such instance involves investigating COVID-19 misinformation on social media platforms like Twitter and Facebook, as well as on news websites and online forums. The overarching objective is to aid stakeholders in promoting accurate information and mitigating the harmful effects of false claims during the pandemic. Iwendi et al. (2022) developed an approach to combating COVID-19-related misinformation by employing Information Fusion to gather real news data from trusted sources and fake news data from social media. Using deep learning models, 39 features were created from multimedia texts to detect fake news, resulting in a substantial improvement in accuracy. The precision, recall, and F1-Measure metrics demonstrate the effectiveness of the models in discerning between real and fake news, outperforming standard machine learning algorithms. This approach holds promise in addressing the challenges posed by misinformation during the pandemic.

### Special Considerations on Arabic NLP for COVID-19

Over the last few years, there have been several attempts to process Arabic content in a variety of applications. For example, Arabic sentiment analysis using a lexicon-based system for Modern Standard Arabic (MSA) applied to “news” was proposed in Abdul-Mageed and Diab (2011). Similar work was reported on modern Arabic in Mourad and Darwish (2013), by using random walks on graphs, while employing Naïve Bayesian and SVM classifiers. An example of using a dataset of Arabic social media content and POS tagging for multi-genre multi-dialect sentiment analysis can be checked in Abdul-Mageed and Diab (2014). Datasets and deep learning models for Arabic text classification were also proposed in Elnagar et al. (2020). Other attempts were also performed to process tweets and Arabic microblogs. A lexicon-based sentiment analyzer for both MSA and Egyptian dialectal Arabic tweets has been developed using an SVM classifier (Heikal et al., 2018). Arabic sentiment analysis of Twitter data related to COVID-19 was presented in Alanazi et al. (2020), with the aim at extracting and ranking the common symptoms discussed among patients on social media. The results were reported from 463 Twitter users who reported being tested positive, with 66% reporting symptoms. Among the symptomatic patients, the top three reported symptoms were fever, headache, and anosmia. Event detection from social media was recently discussed in Ibrahim et al. (2015) using a language-independent Naïve Bayes classification model. The focus was only on specific type of ‘disruptive’ events rather than a generic event detection platform.

Based on the World Health Organization (WHO) definition, an infodemic depicts the use and spread of false or misleading information over any kind of physical or

digital media<sup>1</sup>. An Arabic infodemics study was presented in Shaar et al. (2021). The authors have designed a pilot annotation for English and Arabic organized into seven questions about the input tweet streams. They annotated 504 English and 218 Arabic tweets with a seven-class labeling schema, focusing on the most retweeted ones. They used pre-trained transformers for word embeddings: (i) AraBERT, (ii) FastText, and (iii) BERT, and an SVM classifier. They have argued for the need for a holistic approach to counter the global infodemic related to COVID-19. They stated that the problem is not only in the context of the COVID-19 infodemic, malicious content, and conspiracy theories but also endorsing fake cures, panic, racism, xenophobia, and mistrust in authorities.

Another approach to fake news detection in Arabic is presented in Alsudias and Rayson (2020). The objective was to identify main topics, detect rumors, and predict tweet sources by using k-means clustering and ML classifiers with manual labels on false information. They collected a dataset of tweets related to COVID-19 from December 2019 to April 2020, which contained 1,048,575 unique tweets. They provided a labeled sample of 2000 tweets annotated for false, correct, and unrelated news. Around 60% of the rumors found on Twitter were reported by health professionals and academics, which shows the risk and urgent demand to alert against such fake news.

Analyzing COVID-19 related content in Arabic presents unique challenges in Natural Language Processing (NLP) due to dialectal variations, code-switching, lack of standardization, limited language resources, sentiment analysis and cross-lingual information retrieval (Bahja et al., 2020). Dialectal variations across Arabic-speaking regions necessitate adaptable NLP models capable of handling diverse linguistic forms, while code-switching between Arabic and other languages requires proficiency in recognizing mixed text. Moreover, the lack of standardized COVID-19 terminology in Arabic complicates information extraction, highlighting the need for specialized resources.

## COVID-19 Mining Techniques for Other Types of Data

Although our focus in this survey is to discuss techniques for mining COVID-19 insights from social data, for completeness, we will present in this section other COVID-19 mining techniques related to other types of data, starting from medical imaging to time-series data published by health organizations and online platforms, and finally techniques related to contact tracing. Table 4 illustrates some of the main ongoing research in medical imaging for COVID-19 early screening, while Table 5 presents contributions in contact tracking and time series data analysis.

### Medical Imaging

Early COVID-19 screening through X-ray image classification has been studied in many recent works (Chowdhury et al., 2020; Jain et al., 2020; Zebin & Rezvy, 2020; Turkoglu 2020). Deep learning classification based on X-ray or CT medical imaging and trained on labeled image datasets is the most dominant approach in this

<sup>1</sup> <https://www.who.int/health-topics/infodemic>

**Table 4** Medical image classification during the COVID-19 pandemic

Reference	Purpose of study	Detection technique	Type of analytics	Datasets	Emerging technologies	Evaluation metrics
Soomro et al. (2022)	COVID-19 screening through X-ray image classification	CNN-based image classification	COVID-19 screening	x-ray image datasets	Deep Learning	accuracy, precision, recall, F1 score
Chowdhury et al. (2020)	COVID-19 screening through X-ray image classification	CNN-based image classification	COVID-19 screening	x-ray image datasets	Deep Learning	accuracy, precision, recall, F1 score
Ouchicha et al. (2020)	Classification of chest x-ray images	CNN-based three class classification	classification of normal, viral pneumonia and COVID-19 x-ray images	training with 219 COVID-19, 1341 normal and 1345 viral pneumonia chest x-ray images	Deep Learning	accuracy, precision, recall, F1 score
Tabik et al. (2020)	Building a database for COVID-19 triage systems	Class-inherent transformations (CIT) Network inspired by GANs	quantifying COVID-19 in terms of severity levels, normal, mild, moderate and severe	426 positive and 426 negative PosteroAnterior CXR	Deep learning neural networks	accuracy, precision, recall, F1 score
Umer et al. (2021)	Classification of COVID-19 x-ray images	CNN, image data generator from Keras	Four-class classification, Performance evaluation with VGG16 and AlexNet	10,000 augmented X-ray images	Deep Learning	accuracy, precision, recall, F1-score
Turkoglu (2020)	COVID-19 X-ray image classification	Feature extraction from CNN layers using Relief feature selection algorithm, and SVM classification	classification of COVID-19, normal, and pneumonia cases	6092 X-ray images	Deep Learning and ML	accuracy, precision, recall, F1-score

Table 4 continued

Reference	Purpose of study	Detection technique	Type of analytics	Datasets	Emerging technologies	Evaluation metrics
Zebin and Rezvy (2020)	Transfer learning for classifying COVID-19 chest X-ray images	CycleGAN for image augmentation, transfer learning, CNN	Distinguishing inflammation in lungs due to COVID-19 and Pneumonia from normal cases	673 X-ray images and CT images	Deep Learning, GANs	accuracy, precision, recall, F1-score
Perumal et al. (2020)	COVID-19 CXR classification through transfer learning	Transfer Learning and Haralick features	texture feature extraction for early screening with disease labels	81,176 observations with disease labels	Transfer learning and deep learning	accuracy, loss graphs, precision, recall, F1-score

**Table 5** Contact Tracing (CT) and Time Series (TS) data mining during the COVID-19 pandemic

Reference	Purpose of Study	Detection technique	Category	Type of analytics	Datasets	Emerging technologies	Evaluation metrics
Bradshaw et al. (2021)	Bidirectional contact tracing	hybrid digital tracing, exposure notification, stochastic branching-process model	Contact Tracing	Effect of manual and digital hybrid tracing, benefits of bidirectional tracing	data available here <sup>a</sup>	Digital tracing, exposure notification	Mean effective reproduction number $R_{eff}$
Tran and Nguyen (2021)	Understanding users' decision to use COVID-19 contact-tracing app	Risk-risk tradeoff	Contact Tracing	Convergent validity, PLS-SEM, and Fuzzy-set qualitative comparative	Online survey recruiting participants	Privacy calculus	Convergent validity, cross-validated redundancy index, and consistency and coverage
Chan and Saqib (2021)	Examine if COVID-19 concerns reduce uptake of contact tracing apps	Experimental design	Contact Tracing	One-way ANOVAs, two-way ANOVAs, and logistic binary regression	surveys across France, Australia, and the United States	Hypothetical development	Experimental methodology
Jamieson et al. (2021)	Investigate attitudes toward installing and using contact tracing app	Quantitative and Qualitative analysis	Contact Tracing	Quantitative and Qualitative analysis	Survey of 153 working people and 15 follow-up interviews	Multivariate Regression models	Factor analysis on the correlation matrix of measuring variables
Mokbel et al. (2020)	Privacy-preserving contact tracing through technological facilities	Paradigm shift from personal to facilities, accessibility	Contact Tracing	Focus on unknown contacts, context-aware, privacy-preserving contact tracing	55	Location-based and BLE-based Contact Tracing, Privacy	55



Table 5 continued

Reference	Purpose of Study	Detection technique	Category	Type of analytics	Datasets	Emerging technologies	Evaluation metrics
Luo et al., (2020)	DeepTrack: spatio-temporal monitoring of COVID-19 data	linked mon, ad-hoc, and recommended visualizations, ETL-based data integration	Time series	High-risk area discovery, choropleth map, infection path, similar trend search	real-time COVID-19 data	Interactive visualization	55
Park et al. (2020)	Contact tracing of COVID-19 cases in Korea	indexing confirmed cases, high-risk and non-high-risk groups, track contacts by linking large databases	Contact Tracing	The role of household transmission amid reopening of schools and loosening of social distancing	59,073 contacts of 5,706 COVID-19 index patients	Statistical analysis	55
Ho et al. (2023)	Privacy concerns on mobile contact tracing	Partially anonymizing via polling, random tokenization, private messaging system	Contact Tracing	Studied privacy issues with Singapore's CT app	—	Mobile Tracing, Notions of Privacy	55
Hernandez-Matamoros et al. (2020)	Study the relation between COVID-19 behavior and population in a region	ARIMA - auto-regression, analysis and prediction	Time-series	Evaluated the Auto-Regressive Integrated Moving Average (ARIMA) model for 145 countries distributed over 6 regions	official data of discovered and death cases from ECDC, WHO, JHU, UN, World Bank, etc.	Auto-Regression	RMSE - benchmarking

Table 5 continued

Reference	Purpose of Study	Detection technique	Category	Type of analytics	Datasets	Emerging technologies	Evaluation metrics
Annas et al. (2020)	Stability Analysis of COVID-19 spread in Indonesia	SEIR mathematical generation matrix method	Time-series	Effect of variations in population isolation time on the number of Exposed and Infected cases	Numerical simulated values	Theoretical predictive modeling	benchmarking
Zeroual et al. (2020)	Deep learning methods to forecast the number of new cases and recovered cases	Deep learning using LSTM, BiLSTM, GRU, and VAE methods	Time-series	Forecasting time series of numerical daily discovered and death cases	Numerical simulated values	Deep learning for forecasting	MAE, RMSE, MAPE, EV, and RMSLE

<sup>a</sup> <https://github.com/willbradshaw/covid-bidirectional-tracing>

research field. We, therefore, provide an overview of existing work on deep learning approaches for medical image classification, focusing on COVID-19 detection methods. The authors in Ouchicha et al. (2020) proposed a methodology for early screening of COVID-19 cases based on chest x-ray images using CNN-based three-class classification: 1) normal, 2) viral pneumonia, and 3) COVID-19. They trained their model with 219 COVID-19 X-ray images, 1341 normal and 1345 viral pneumonia chest X-ray images and evaluated the performance based on accuracy, precision, recall, and F1 score. Promising results of accuracy of 96.69% were reported on the three class classifications. However, most of the approaches in this domain present an issue related to the small training data size, which may impact the scalability of such models for real-life diagnosis. Other approaches proposed to augment training data by multiplying the data size using techniques, such as Generative Adversarial Neural Networks (GANs) and Keras image data generator (Tabik et al., 2020; Umer et al., 2021; Zebin & Rezvy, 2020).

Tabik et al. (2020) proposed building a database for COVID-19 triage systems using class-inherent transformations (CiT) network inspired by GANs. Umer et al. (2021) used CNN classification to quantify COVID-19 cases in terms of severity levels: normal, mild, moderate, and severe. They trained their model on 426 positive and 426 negative chest X-ray images and a generated dataset of 10,000 images using the ImageDataGenerator class from Keras. Radiomic features and ML algorithms were combined in another approach for the early detection of COVID-19 and distinction from other types of viral/bacterial chest infections (Tamal et al., 2021). Similarly, Zebin and Rezvy (2020) used transfer learning for classifying COVID-19 chest X-ray images and CycleGAN for image augmentation. They aimed at distinguishing inflammation in the lungs due to COVID-19 and Pneumonia from normal cases based on labeled 673 X-ray and CT images. Another COVID-19 X-ray image classification approach used feature extraction from CNN layers using the Relief feature selection algorithm and SVM classification (Turkoglu, 2020). Perumal et al. (2020) presented a COVID-19 CXR classification method through transfer learning and extracting Haralick features. They claim texture feature extraction can be very helpful for early screening. Their trained model contains 81,176 observations with disease labels. Janarthanan et al. (2021) present a study on how artificial intelligence and media imaging can be utilized to diagnose COVID-19 patients. The authors extracted data from various research reports, articles, and WHO guidelines to identify the disease's diagnosis, treatment strategies, and outcomes.

Alelyani et al. (2021) provides an evaluation study on the impact of the COVID-19 pandemic on medical imaging. The idea was to study how imaging volumes and imaging types in radiology are affected by COVID-19 in various locations. The authors utilized images between 2019 and 2020 from different hospitals that include cases related to outpatient, inpatient, and emergency departments. Such data was compared using t-tests. The results show that there was a decline observed in outpatient departments by 76% and emergency departments by 25%. Moreover, there was a decrease in nuclear medicine, ultrasound, MRI, and mammography by 100%, 76%, 74%, and 66%, respectively. Born et al. (2021) offer a systematic review on the use of AI in imaging for COVID-19. The Authors have covered 463 papers published on AI for imaging-related studies. Their findings showed a significant disparity between clini-

cal and AI communities in focusing on both imaging modalities. Furthermore, most of the research was found to be lacking concerning potential use in clinical practice. Furthermore, the authors in Aytac et al. (2022) suggest that applying an adaptive momentum rate for image classification would reduce classification error and increase accuracy.

The authors in Quak et al. (2021) have studied the relationship between gender disparity in medical imaging research and the COVID-19 pandemic. The goal was to investigate the impact of female physicians' research in medical imaging on scientific publications. As a result, the researchers gathered information from 50 medical imaging papers published between March and May 2020. The result shows that there is the gender imbalance in the first and last authorship for articles submitted to the top 50 medical imaging journals. Rehouma et al. (2021) provide a comprehensive review on the use of machine learning models in COVID-19 detection. 62 papers based on deep learning algorithms were selected for analysis. The authors illustrated that convolutional neural networks have been widely used for image segmentation and classification to detect patients with COVID-19.

## Contact Tracing

Contact tracing is another very important field of research that has witnessed a big adoption and government support by developing mobile contact tracing applications that help in tracking of confirmed cases and in reducing the spread of the disease (Ahmed et al., 2020). Many applications have been published to reduce the fast COVID-19 spread. Nonetheless, this approach has failed to some extent in achieving its purpose for many reasons (Dar et al., 2020). A survey on existing applications can be found in Ahmed et al. (2020). Our objective in this section is to highlight new approaches and discuss techniques that try to cope with the issues encountered in such mobile tracing applications.

Contact tracing of COVID-19 cases in Korea was studied in Park et al. (2020). They proposed indexing confirmed cases, high-risk and non-high-risk groups, and tracking contacts by linking to large databases (59,073 contacts and 5,706 COVID-19 indexed patients). They aimed to highlight the role of household transmission amid the reopening of schools and the loosening of social distancing. Bradshaw et al. (2021) introduced a hybrid bidirectional contact tracing with a digital exposure notification based on stochastic branching-process modeling. The effect of manual and digital hybrid tracing to identify infectors and their infectees and the benefits of bidirectional tracing were investigated. Another approach for contact tracing using indoor trajectories of moving users was introduced Alarabi et al. (2021), which considers social distancing and the exposure period to find potential infectees.

Tran and Nguyen (2021) investigated the risk-risk tradeoff model based on the privacy calculus theory and the risk-risk tradeoff notion to understand better COVID-19 contact-tracing app users' risk minimization decisions. According to their findings, users participate in a health risk-privacy risk tradeoff while considering and opting to use the applications. As a result, their study contributes to the field of privacy calculus theory research and argues for a balanced management solution to this trade-off challenge. The authors in Chan and Saqib (2021) conducted three experiments

in France, Australia, and the United States to see if key COVID-19 issues, which should raise worries about personal and public health, do raise privacy concerns, lowering the use of contact tracing applications. They discovered that notable COVID-19 concerns reduce intentions to use contact tracing applications using an experimental design in which individuals were randomly assigned to either a disease concern or a control condition. The mediation findings show that higher privacy values explain the lesser willingness. Jamieson et al. (2021) evaluated attitudes on downloading and utilizing contact tracking apps and how they linked to respondents' everyday lives, work patterns, and overall sentiments about the epidemic using a survey of 153 working individuals and 15 follow-up interviews. They discovered that the incentives for downloading the app differed from those for continued use. They looked at how people navigated ambiguous behavior norms during the epidemic and considered personal risks while determining whether to use contact tracking apps.

The main challenge discussed in mobile contact tracing is the privacy concerns concerning revealing users' detailed movements and contacts in real-time (Mokbel et al., 2020). Privacy-preserving Contact tracing through technological facilities was recently proposed in Mokbel et al. (2020). It suggests a paradigm shift from personal tracking through GPS or BLE-based techniques to large infrastructures and facilities, thus achieving better accessibility to elderly people and less exposure to users' private data. Privacy concerns on mobile contact tracing has been also investigated in Cho et al. (2020). They discussed different privacy-aware methods with a use case on Singapore's contact tracing app using partial anonymization via polling, random tokenization, and private messaging systems.

### Time Series Data Mining Techniques

DeepTrack is a real-time dashboard for spatio-temporal monitoring of COVID-19 data (Luo et al., 2020). Different types of interactive visual analytics were used, such as choropleth maps, linked common, ad-hoc, and recommended visualizations. Other systems apply ETL-based data integration and generate analysis related to high-risk area discovery, tracking infection path, and similar trend search in real-time (Leung et al., 2022).

The analysis of time series COVID-19 health data through 6 geographic regions was presented in Hernandez-Matamoros et al. (2020). The authors introduced a relationship model between countries in the same geographical region to predict the spread of the virus. They evaluated their algorithm using the Auto-Regressive Integrated Moving Average (ARIMA) model for 145 countries distributed over six regions, with parameters that include population per 1 million people, the number of cases, and polynomial functions. Their results show the potential to create other models to predict the pandemic behavior using other variables, such as humidity, climate, and culture. They collected data from the European Centre for Disease Prevention and Control (ECDC), the WHO, Johns Hopkins, the United Nations, the World Bank, the Global Burden of Disease, and the Blavatnik School of Government. In a different approach, the social impact of the COVID-19 pandemic on the employment promotion policies for graduate students in China was studied in Chen et al. (2021).

A statistical model is proposed in Dash et al. (2021) to forecast the outbreak of COVID-19 such as future peak dates and change points in the growth of the pandemic by analyzing time series data of new cases. Also, By analyzing time series of suicide data from several countries, the authors in Pirkis et al. (2021) found that the number of suicide cases related to COVID-19 remained mostly unchanged or declined in the early months of the pandemic as compared to the expected number of suicide numbers due to the pandemic. ARIMA modeling to forecast the expected daily number of COVID-19 cases in Saudi Arabia was also explored in Alzahrani et al. (2020). The model is tested on 7668 new cases per day and over 127,129 cumulative daily cases in four weeks. The forecasting results showed the trend in Saudi Arabia compared to the prediction of new cases from the official website of the Saudi Ministry of Health. The prediction evaluation of daily discovered and death cases was performed using RMSE, MAPE, RMSRE values, and the highest R2 values.

Stability Analysis of the COVID-19 spread in Indonesia was studied in Annas et al. (2020), simulating the SEIR mathematical model on COVID-19 data. They constructed the SEIR model by considering vaccination and isolation factors as model parameters and used the generation matrix method for data analysis. A comparative study of five deep learning methods to forecast the number of new cases and recovered cases from six countries (Zeroual et al., 2020). Recurrent Neural Network (RNN), Long short-term memory (LSTM), Bidirectional LSTM (BiLSTM), Gated recurrent units (GRUs), and Variational AutoEncoder (VAE) algorithms were developed to demonstrate the promising potential of deep learning models in forecasting COVID-19 cases. VAE achieved better forecasting performance of new and recovered cases than all other models. Datasets were made publicly available by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. The five DL forecasting methods were assessed using MAE, RMSE, MAPE, EV, and RMSLE metrics for each country.

Deep learning forecast models are also proposed in Shahid et al. (2020) for time series prediction of confirmed cases, deaths and recoveries in ten major countries affected due to COVID-19. Developed models include ARIMA, support vector regression (SVR), LSTM, and Bi-LSTM. Bi-LSTM model outperforms in terms of endorsed indices and can be exploited for pandemic prediction for better planning and management. Predictive modeling to improve the current time-series forecasting algorithms was studied in Zivkovic et al. (2021). Katris (2021) proposed a technique to produce a time series based on statistical data to track the spread of the COVID-19 pandemic. The technique used Multivariate Adaptive Regression Splines and Feed-Forward Artificial Neural Networks to analyze and predict the spread of the disease. A hybrid approach (Hybridized CESBAS-ANFIS) combining machine learning and nature-inspired algorithms was claimed to perform better than other approaches. Forecasting the COVID-19 spread was also studied in Desai (2021), where a multivariate CNN trained on test positivity data combined with news sentiments derived from IBM Watson Discovery News. Their model shows a spread prediction accuracy that is higher than the baseline Bayesian-based SEIRD model. The authors in Kuo et al. (2021) proposed a hybrid prediction approach based on county-level demographic, environmental, and mobility data. Multiple machine learning techniques and a hybrid framework were implemented to discover high infections on weekends when mobility increases and the effect of long and short lock-downs.

Eight machine learning algorithms were employed: (1) elastic net (EN) model, (2) principal components regression (PCR) model, (3) partial least squares regression (PLSR) model, (4) k-nearest neighbors regression (KNN) model, (5) regression tree (RT) model, (6) random forest (RF) model, and (7) gradient boosted tree models (GBM), and (8) a 2-layer artificial neural network (ANN) model. The approach was evaluated on data extracted from official health platforms for new cases and population mobility data. The results demonstrate better daily case prediction for random forest (RF), gradient-boosted tree models (GBM), and a 2-layer artificial neural network (ANN) model. At the same time, elastic net (EN) and GBM predicted well for cumulative cases.

## Evaluation Metrics

This section emphasizes the need for evaluation metrics and benchmarks in each research field discussed above to achieve a deeper understanding and assessment of the proposed solutions. These metrics are domain-dependent. A summary of evaluation metrics per research field is illustrated in Table 6. Note that most of the available studies on social data mining lack rigorous performance evaluation since they focus on producing an analysis of social data concerning topic modeling and the spread of misinformation. On the other hand, all approaches in medical image classification adopt the accuracy, precision, recall, and F1 scores to assess the classification performance. Contact tracing performance is assessed by considering the user uptake and adherence, measuring the quarantine infectious people as accurately as possible, the quick and

**Table 6** Evaluation metrics per research field

Data Source	Articles	Metrics
Social media mining	Koh and Liew (2020), Mutlu et al. (2020), Abd-Alrazaq et al. (2020), Gencoglu (2020), Alsudias and Rayson (2020)	Accuracy, F1, Recall, Precision, highest/lowest means, TF-IDF
Medical imaging	Chowdhury et al. (2020), Ouchicha et al. (2020), Tabik et al. (2020), Umer et al. (2021), Turkoglu (2020), Zebin and Rezvy (2020), Perumal et al. (2020)	Accuracy, Loss graphs, Precision, Recall, F1-score
Contact tracing	Colizza et al. (2021), Mokbel et al. (2020), Cho et al. (2020), Park et al. (2020), Vecino-Ortiz et al. (2021), Ahmed et al. (2020)	Mean effective reproduction number $R_{eff}$ , standard error from the effect size and p-value
Time series data	Luo et al. (2020), Hernandez-Matamoros et al. (2020), Annas et al. (2020), Zeroual et al. (2020), Shahid et al. (2020), Zivkovic et al. (2021), Kuo et al. (2021)	MAE, RMSE, MAPE, EV, and RMSLE

secure notification, and the ability to evaluate effectiveness transparently (Braithwaite et al., 2020; Colizza et al., 2021). However, most approaches do not consider such advanced metrics in mobile contact tracing. Time series data is well well-established research domain. Thus, evaluation metrics are based on assessing mathematical models and foundations such as MAE, RMSE, MAPE, EV, and RMSLE metrics to measure the forecast accuracy and bias (Vandeput, 2021).

## Analytical Perspectives

COVID-19 has recently been the subject of a slew of analytical research. For example, the authors in Chernozhukov et al. (2021) assess the complex effects of the different US-state policies on reported COVID-19 cases and deaths and Google Mobility Reports for social-distancing activity. In this study, the voluntary reaction of individuals to news on transmission threats was considered in a causal structural model context. A review of their findings suggests that policy and risk communication knowledge are key determinants for the cases of COVID-19 (death rate). Also, it indicates that a shift in policy represents a significant proportion of the improvements found in social distancing behavior. In Istanbul, a panel data analysis was presented by Shakibaei et al. (2021).

A framework developed in Agarwal et al. (2020) tries to explain as part of the dialogue the various incidents that may occur during a pandemic based on social media exchanges. Various COVID-19 disorder strategies have been implemented, including identifying outbreaks, monitoring viral propagation, diagnosis and treatment, the detection of vaccines, and drug research. Other research challenges, including data safety, inconsistency in pattern, control, and transparency of diseases, and the difference between the symptoms of COVID-19 and non-COVID-19, were studied in Bhattacharya et al. (2021). Despite the good results, successful DL processing of COVID-19 medical images still takes considerable time and effort and close activity among various parties in government, industry, and academia. The authors in Grasselli et al. (2020) proposed a framework for predicting up to two weeks forecast for utilization and availability of ICU beds during the COVID-19 pandemic. The framework uses an ensemble approach that combines autoregressive, artificial neural networks and a compartment model. The system was tested on a Chile dataset, achieving a mean error of 4% for the first week and 9% for the second week. Results showed the ensemble approach performed better than individual models for handling different scenarios.

The transmission model of the SARS-CoV-2 virus was used to develop a new dynamical model based on flow networks (López & Čukić, 2021). The proposed model was developed using ‘SEQIJR model’, which can detect SAR-COV-1 network flow. The network analysis enables the transport flow to be defined as a linear programming issue while some functions within the device limitations need to be optimized. The scalability and adaptability of the system to various sub-populations is a further advantage. However, the lack of adaptability to a specific region and the inclusion of elderly people nodes into the network are major deficiencies of the model.



## Social Media Analysis

Social media has been a great source of data for analysis. As the news of COVID-19 keeps spreading globally, some countries have yet to believe the existence of the deadly disease. The fake news spreading phenomenon has been studied in Apuke and Omar (2021) to identify how six different variables affect the outcome of fake news spreading. The findings suggest generosity was the most important factor in predicting COVID-19 fake news distribution. However, the research failed to assess the impact of fake media sharing on cultural context, age, and gender because the study was applied to Nigeria, a country with multiple cultural and ethnic groups. Also, research was carried out among residents in Wuhan by Zhong et al. (2021), which was the origin of the COVID-19 outbreak virus. The authors examined how Wuhan residents processed the health information on social media and how their use of social media could reveal a risk to mental health at the highest rate of the Wuhan-19 Outbreak. Their study's results can help explain the potential connections between the use of social media and the mental distress experienced by individuals in the public health crisis. Furthermore, the study also provides insights into the mechanism of health training and public reaction to pandemics for a deeper understanding. However, the study does not address how to design potential interventions and health policies that alleviate the impact on mental health during or after the COVID-19 crisis.

In addition, from the perspectives of the pandemic in Atlanta, San Francisco, and Washington DC, it offers practical lessons for city governments and highlights the theoretical value of focusing on public relations methods through government. As a result, Zeemering (2021) carried out an exploratory investigation based on Functional fragmentation in city hall. The data is collected from city agency Twitter accounts and key informant interviews to validate the significance of fragmentation for core organization, as well as public outreach. However, the research was undertaken during the early stages of the US response to COVID-19. Similarly, Shahi et al. (2021) conducted an investigative study on COVID-19 misinformation from Twitter using alternative and complementary approaches to conduct an exploratory analysis into the Twitter accounts behind COVID-19 misinformation, COVID-19 misinformation dissemination on Twitter and false assertion material on COVID-19 circulating on Twitter. They focused on the decisions of experienced fact-checking organizations that track every argument manually.

From another perspective, perinatal health promotion content over Facebook and other websites during the pandemic was investigated in Durowaye et al. (2022). The authors concluded that although diverse topics related to healthy pregnancy during COVID-19 were covered on social media, many gaps were found in spreading the severity or risks during pregnancy and fighting misinformation. In Teng et al. (2022), they have analyzed 43K YouTube comments in order to infer the reasons behind vaccine hesitancy among users. In particular, concerns related to safety and potential side effects were raised by users, in addition to a lack of trust in authorities' decisions and pharmaceutical companies. The authors suggested that anti-vaccination activists over social media have spread a lot of misinformation, which led to an amplification in vaccine hesitancy.

## Contact Tracing from an Analytical Perspective

To understand the effect of social distancing measures on Brazilian MSM and transgender/non-binary lives, a web-based survey was conducted by Torres et al. (2020). The authors used personal individual lives, pre-exposure Prophylaxis/Antiretroviral Therapy access, and sexual behavior for their research. Both PrEP and ART devices are used for the determination of people with HIV in testing that is positive/negative. These methods also help us to analyze the factors linked to the failure to sustain social distance. Similarly, the influence of social distancing procedures of COVID-19 was examined by Castex et al. (2020), using a cross-country variation in dimensions of the socio-economic, regional, environmental, and health systems. Moreover, with demographic densities, country surface area, the working rate, and the proportion of elderly in the population decreasing and rising per capita GDP and health spending in the proceeding, the efficacy of procedures prescribing school closures and jobs. According to the authors, these results are by cross-country human mobility data reinforcement. This is because policies are similar across countries, while country characteristics vary substantially. Privacy is still the biggest concern when tracing people's location or interactions, as discussed in Liu et al. (2023). The authors suggest that a privacy-preserving efficient tracing solution can be developed by combining some intrinsic properties of blockchain, such as anonymity, decentralization, and traceability.

## Emerging Technologies for COVID-19 Data Analytics

The analysis of Evolving Technologies used for treating and Diagnosing COVID-19 was discussed by Vafea et al. (2020). This analysis outlines the new technology used in the COVID-19 research, diagnosis, and treatment. Key fields of focus include artificial intelligence, Big Data, and the Internet of Things, the relevance of mathematical prediction models, the use of Community screening technology and nanotechnology, the use of telemedicine to manage new demands, and the potential of robotics and other technologies. Table 7 summarizes the emerging technologies for COVID-19 applications.

## Impact on Social Behavior

Efforts to predict the spread of COVID-19 provide valuable insight to fight COVID-19. The social behavior of travelers in Istanbul during the COVID-19 pandemic was presented in Shakibaei et al. (2021). The study examines the impacts of the pandemic on travel behavior based on descriptive research using specific tri-wave evidence. The results of this study cause the Turkish government to take some action on the conduct of individuals traveling in Istanbul and to discern various trip uses, such as homework, social/recreational/leisure (SRL), and shopping. In another important study, the impact of CoronaVirus on Education in England and the impact of online learning on parents, teachers, and students between 11 to 15 years old was investigated (Connor et al., 2022). Interestingly, among 329 parents/carers and 117 teachers, one-third of teachers

**Table 7** summary of the emerging technologies for COVID-19 applications

Emerging technology	COVID-19 applications	Ref.
Artificial intelligence	Classifying people based on their network	
	Body temperature checking	Mbunge (2020)
	Detect people who infected with COVID-19 virus	Wang et al. (2020)
Internet of things	Real-time monitoring and responding via 'chatbox', authorities will have full control over access COVID-19 data for contact-persons in real-time.	Mbunge (2020)
	Availability of the contact-persons on the cloud	Verbeek et al. (2020)
	COVID-19 patients in quarantine centers and/or self-isolation are monitored remotely	Kang et al. (2020)
Big data	System enabling real-time access to national health information systems and storage of huge data for additional analysis. COVID-19 monitoring and resource allocation.	Zhou et al. (2020)
	Raise awareness, education and communication.	Tan & Lin (2023)
	Support video conferencing activities.	Sadler et al. (2020)
Blockchain technology	Authenticate and validate COVID-19 patients before the provision of counseling services	Mbunge (2020)
	In a peer-to-peer architecture, encrypt health data as it passes from one node to the next	Yasaka et al. (2020)
5G technology	To assure real-time data access, increase bandwidth size, perceived quality of service, and data transfer rate.	Liang et al. (2020)
	Facilitates tele-consultation between healthcare professionals and COVID-19 patients and contact-persons.	Mbunge (2020)

and around half of parents reported below-average well-being due to issues related to access to resources, confidence in online teaching, etc. Parents also revealed concerns about their children's mental health and lack of access to electronic devices.

Considering the following factors: molecular, environmental, and social factors, the effect of the COVID-19 pandemic on children and adolescents' mental health was studied by de Figueiredo et al. (2021). These considerations were taken into

account because the sudden separation from the classroom, social life, and sports in the open air has significantly influenced children and young people. Some have endured increasing domestic abuse. However, the paper seeks to address the need for supervision and treatment for these people and to alert public health and government agencies.

### **Impact on Businesses and Economy**

The consequences of (the COVID-19) pandemic have also produced a profound impact on business and the economy. For instance, Silva et al. (2020) evaluated how econometrics, machine learning models, and ensemble methods can be used to predict new COVID-19 cases. In econometrics, the study used ARIMA and SARIMA econometrics models. For machine learning models, AdaBoost and GBR models were evaluated. Moreover, ensemble methods were also evaluated. The study evaluated these models on the Brazil, South Korea, China, and Italy datasets, using features such as total number of cases, deaths, new cases, new deaths in the day, and recovered patients. The results showed that no single model gave better predictions in all datasets. However, the ensemble of machine learning and econometrics showed great potential. This is because machine learning models perform poorly with less data but can be compensated by using ensemble methods. The relationship between employment conditions and protective measures among low-income US workers during the pandemic has been studied in Capasso et al. (2022). Findings suggest that essential workers struggled with variable income or income loss, unpaid sick leaves, and others have suffered from food insecurity.

### **Sentiment Analysis**

The initial impacts on the US stock market on COVID-19 Sentiment using Big Data was also considered by Lee (2020). This research examined the association between the COVID-19 sentiment and 11 selected United States (US) stock market sector indices between 21st January 2020 to 20th May 2020 on coronavirus-related searches, using the Daily News Sentiment Index (DNSI) and Google Trends data. Almost no use of tweeter data was made of DNSI or Google Trends during intensive studies on sentiment analysis to forecast the stock market movement. Moreover, this analysis explores the difference in US business forecasting shifts in DNSI by estimating a time-series model of regression with excess industry returns as the dependent variable. The authors in Pham et al. (2022), study the impact of the COVID-19 pandemic on the financial markets by studying US ex-President Trump's tweets in order to infer industry-level reactions based on his tone of speech during the pandemic. The relationship and statistical correlation to 49 industries were discussed by analyzing the sentiments of 2574 tweets from Trump's Twitter account.

Similarly, by monitoring COVID-19-related Twitter updates, the authors in Naseem et al. (2021) discussed the topic of Twitter sentiment using benchmark sentiment analy-

sis approaches. Their research findings show that the population favored a lockdown in February, but their view shifted to mid-March. While the reason for the shift in feeling is unclear, misinformation is spread across social media and the need for a proactive and agile presence in public health is therefore necessary for the fight against the spread of false news. Also, the authors have published a wide-scale COVID-19 data collection for the study of emotion, which is freely accessible. In Italy, an Information Management System to Detect and Monitor Italian Tweets Relevant Topics During the COVID-19 event was developed by De Santis et al. (2020). To this end, a methodological paradigm based on a biological metaphor has been experimented with, which can monitor new words and evolving concepts over time, beginning with a real-world Tweets dataset gathered during the lockdown. The technique was a driver for creating an ongoing Twitter monitoring scheme expressly designed to retrieve the Italian language's buzzwords and subjects. Besides, the proposed system can discover the newest conditions for socio-political activities in an uncontrolled way, which is highly emphasized, even for words that are often and continuously used, such as the names of leading prime ministers. It is also generally used to identify and track issues arising from socially important events in feeds of Twitter messages written in either language.

From a different perspective, sentiment analysis of users' tweets concerning COVID-19 vaccines, such as Pfizer, Moderna, and Sinopharm, was presented in Mush-taq et al. (2022). Users' sentiments on vaccines in general and then on each vaccine with its geographical distribution were reported. Temporal tracking of peak discussion times for specific vaccines and their spatial whereabouts. Overall, sentiments on related topics have changed over space and time, and the overview given can help policymakers adjust their policies in order to enhance their vaccination program acceptance.

## Prediction

Several research studies have investigated the likelihood of predicting occurrences and trends of the COVID-19 pandemic. Elsheikh et al. (2021) proposed using a deep learning model using a long-short-term memory model for predicting the number of total confirmed cases, recovered cases, and deaths due to COVID-19 in Saudi Arabia. The proposed model are also tested for other countries as well for verification purposes, including Brazil, India, South Africa, Spain, and the USA. The system utilized the optimal hidden value and learning rate to achieve better results, which were 100 and 0.005, respectively. The system could predict results up to 1 week, which is far better with baseline systems tested against, including NARANN and ARIMA. The system also used several evaluation metrics for testing the results, including Root mean square error, coefficient of determination, mean absolute error, efficiency coefficient, overall index, coefficient of variation, and coefficient of residual mass. In the coefficient of determination, which highlights the correlation of predicted results vs. actual results (with a score between 0 and 1), the system achieved 0.976 for total cases and 0.944 for total deaths.

Progressively, Adly et al. (2020) have introduced an automated tool for computer Tomography (CT) image analysis to deal with the outbreak of COVID-19 using a deep learning approach. The idea is to detect, track and quantify COVID-19, which can distinguish between patients infected with COVID-19 and those who are not. The study used a variety of worldwide databases, including disease-infected areas in China. Chieragato et al. (2022) have proposed a severity predictive model to classify ICU from non-ICU patients based on CT images, a 3D CNN for feature extraction, and a CatBoost classifier. The authors suggest that integrating heterogeneous features and a better interpretability of models would greatly enhance the prediction of such complex tasks. From a different perspective, the authors in Kim et al. (2022) have studied the development of quarantine-related programs and the level of adherence to such distancing rules in different communities. The aim was to predict the influential factors and norms on how individuals' compliance with quarantine rules may impact the next wave of COVID-19 spreading.

## Other Perspectives

*Healthcare Infrastructure and Resource Allocation:* Analyze how data mining techniques can be used to optimize the allocation of healthcare resources such as hospital beds, ventilators, and medical personnel based on COVID-19 case data and patient demographics (Arunmozhi et al., 2022).

*Vaccine Distribution and Effectiveness:* Discuss how data mining can aid in the distribution and monitoring of COVID-19 vaccines, including tracking vaccination rates, identifying vaccination hotspots, and assessing vaccine effectiveness through real-world data analysis (Sun et al., 2021).

*Epidemiological Modeling and Forecasting:* Explore the use of data mining techniques in epidemiological modeling to predict the spread of COVID-19, estimate infection rates, and evaluate the effectiveness of intervention strategies such as lockdowns and social distancing measures (Namasudra et al., 2023).

*Public Health Interventions and Policy Evaluation:* Examine how data mining can inform public health interventions and policy decisions, including evaluating the impact of various interventions on disease transmission, healthcare outcomes, and socio-economic indicators.

*Community Engagement and Behavioral Interventions:* Discuss strategies for leveraging data mining to promote community engagement, encourage adherence to public health guidelines, and design targeted behavioral interventions to mitigate the spread of COVID-19.

*Ethical and Privacy Considerations:* Address ethical and privacy concerns associated with COVID-19 data mining, including issues related to data security, informed consent, and the responsible use of sensitive personal information in research and decision-making processes (Anshari et al., 2023).

*Long-Term Socio-Economic Impacts and Recovery Strategies:* Analyze the long-term socio-economic impacts of the COVID-19 pandemic and explore data-driven strategies

for economic recovery, workforce reintegration, and rebuilding resilient communities in the post-pandemic era.

## Opportunities and Challenges

The spread of COVID-19 has created opportunities and challenges for analyzing available datasets such as medical images and tweets, to fight the pandemic. In this section, we highlighted some of the opportunities and challenges in research related to the COVID-19 pandemic from social Media, medical image, and contract tracing perspectives.

### Opportunities

The increased use of artificial intelligence-based techniques will enable social media data to be analyzed in real time. Such analysis provides an opportunity to track changing public sentiments concerning the COVID-19 pandemic and proactively communicate with the public (Hussain et al., 2021). Moreover, there is a pressing demand to identify COVID-19 outbreak rumors, hoaxes, and misinformation on social media, which causes panic among the public. In addition, with the vaccine rolled out, we need to fully understand public sentiments and address the concerns of vaccine skeptics (Hussain et al., 2021). A recent study on data mining algorithms that can be combined with epidemiological prediction models was presented in Cortés-Martínez et al. (2022). The authors consider that such an integration would help develop more accurate prognosis tools for better managing and tracking of viral diseases.

Safdari et al. (2021) have reviewed the most favorite data mining techniques to fight pandemics, such as NLP, for revealing disease characteristics. The authors in Abdalla et al. (2023) suggest that knowledge discovery methods can help infer unknown disease dimensions during the pandemic. A similar study reveals that 90% of techniques apply highly accurate supervised learning for classification or prediction tasks in the epidemiology discipline (Ghosh and Das, 2022).

Furthermore, social engagement among individual users and communities over social media applications is an essential research topic, as it may help in the development of more efficient epidemic models that account for social behavior, as well as more successful and targeted crisis communication tactics (Cinelli et al., 2020). During the COVID-19 epidemic, we will be able to find a significant incidence of mental health disorders that is positively correlated with frequent social media exposure (Gao et al., 2020). Social media can provide opportunities to disseminate and receive details about patients, clinicians, and scientists. Contact tracing apps have raised many concerns about their purpose, privacy breaches, how they operate, authority sponsorship, and the willingness to use such a technology (Abuhammad et al., 2020).

Murphy et al. (2020) identified some opportunities to enhance cognitive behavior therapy during COVID-19. In their study, they found that a potential solution to address the problem of fears of infection and the effects of social isolation is to deliver enhanced cognitive behavior therapy, an evidence-based treatment.

## Challenges

The COVID-19 pandemic has generated enormous data on the spread of the virus, its impact on society, and the response of governments and healthcare systems. Data mining techniques have played a critical role in analyzing this data to gain insights and inform decision-making. While different types of vaccines and booster shots are available nowadays, the spread of the virus has not stopped (Yih et al., 2023). It seems more time is needed to reach herd immunity worldwide, and it is unclear how long newly generated COVID variants can resist or bypass developed vaccine protection (Windsor et al., 2022). Table 8 summarizes some of the existing challenges and opportunities of COVID-19 data analysis. Additionally, there are prospects for leveraging data mining solutions to overcome these challenges effectively. Here are some lessons learned from a data mining perspective:

- **Continuous Virus Spread:** Despite the availability of vaccines and booster shots, the spread of the virus persists. Achieving global herd immunity remains a challenge, compounded by the emergence of new COVID variants that may evade vaccine protection.
- **Real-time Data Utilization:** While real-time data is crucial for decision-making, there are challenges in processing and analyzing large volumes of data in real-time. This requires robust data mining infrastructure and algorithms capable of handling streaming data efficiently.
- **Data Sharing and Collaboration:** While data sharing and collaboration are essential, there are barriers to sharing data across borders and organizations, including privacy concerns and regulatory restrictions. Overcoming these barriers requires international cooperation and the development of standardized data sharing protocols.
- **Predictive Analytics Accuracy:** While predictive analytics has been instrumental in forecasting the spread of the virus and predicting healthcare resource demand, there are challenges in developing accurate and reliable predictive models. This necessitates the refinement of modeling techniques and the incorporation of diverse data sources for improved model performance.
- **Data Quality Assurance:** Ensuring data quality is critical for the reliability of analytical insights. Challenges such as data incompleteness, inconsistency, and bias can affect the accuracy of data mining models. Addressing these challenges requires robust data quality assurance processes and the implementation of data cleansing and normalization techniques.
- **Ethical Considerations:** The use of personal data in data mining raises ethical considerations related to privacy, fairness, and transparency. There is a need for ethical guidelines and regulatory frameworks to govern the ethical use of data mining techniques in the context of the pandemic.

Although many studies have been conducted in the last three years on the COVID-19 pandemic, research on social media, contact tracing, and the impact on economics remains in its early stages. Moreover, new challenges emerge from different aspects (Ajaz et al., 2022). Several challenges have been addressed. However, there still is room for improvement. The authors in Ajaz et al. (2022) suggest that COVID-19 can



**Table 8** Challenges and opportunities per research field

Category	Challenges	Opportunities
Social data mining	<ul style="list-style-type: none"> <li>– Collection of sufficient COVID-19 data/or big data using social media platforms.</li> <li>– Due to the sheer size of social media data, it's not possible to obtain a properly annotated dataset to train the algorithms</li> </ul>	<ul style="list-style-type: none"> <li>– Content consumption and social media platforms have different social dynamics.</li> <li>– Positive association with frequent social media exposure during the COVID-19 outbreak. Patients, clinicians, and scientists will benefit from receiving and disseminating pertinent information. Cuello-Garcia et al. (2020).</li> <li>– A multidisciplinary approach with better outreach and dissemination is optimal. Cuello-Garcia et al. (2020)</li> </ul>
Contact tracing	<ul style="list-style-type: none"> <li>– Dealing with privacy breaches in data containing personal information.</li> <li>– Ethical Issues, i.e., privacy and informed consent (Abuhammad et al., 2020)</li> <li>– Concerns about lack of information about COVID-19 tracing contact application (Lucivero et al., 2020)</li> </ul>	<ul style="list-style-type: none"> <li>– There are no common criteria in place to assure the quality of these tracing technologies and their uses in terms of ethical considerations. Abuhammad et al. (2020).</li> <li>– Tele dermatolog (Gupta et al., 2020)</li> </ul>
Medical imaging	<ul style="list-style-type: none"> <li>– Limited amount of labelled medical images (Zebin &amp; Rezvy, 2020; Ilyas et al., 2020)</li> <li>– In the medical setting, any deep learning-based solution can be used effectively. Singh et al. (2020)</li> <li>– Larger and more heterogeneous databases are needed Luz et al. (2021)</li> <li>– Scarcity of x-ray images Mahmud et al. (2020)</li> </ul>	<ul style="list-style-type: none"> <li>– Utilizing multi-class disease segmentation and in depth analysis (Ilyas et al., 2020)</li> <li>– Early diagnosis of the susceptible patient (Minaee et al., 2020)</li> <li>– Lung Ultrasound (Schultz et al., 2020)</li> </ul>

be controlled using IoT technology and machine learning techniques. A multi-layered architecture of IoT solution has been proposed, where unmanned aerial vehicle (UAV) applications can be used for privacy-preserving contact tracing.

In addition, sentiment analysis experiments were extensively checked over 10 years (1st January to 30th June 2020) in the presence of infectious conditions, infections, epidemics, and pandemics (Alamoodi et al., 2020). The inspiration behind this research was COVID-19's wide distribution. However, COVID-19 remains ambiguous as an infectious disease, as its literature and cases massively proliferate; thus, it is almost difficult to track up-to-date infection. Besides, only after the pandemic stops can correct details be gathered. Further research should concentrate on the role of social media and the study of feelings during a related event.

Because of human tolls and economic implications, COVID-19 has inflicted unparalleled disruption on the global economy. It presented taxpayers and politicians with a larger threat to minimize the impact of this pandemic (Padhan & Prabheesh, 2021). The economic impact of COVID-19 was highlighted in this report, and policy alternatives were highlighted to minimize their impacts. Their study concludes that monetary, macro, and fiscal policy separately contribute to mitigating effects; in the post-pandemic cycle, the combined trio may be more successful. It is also essential that the consequences of COVID-19 be reduced by cooperation between the 'trio' policies, i.e., monetary, macro-prudential, and fiscal policies. Other challenges require a great deal of attention. These challenges include the deactivation of mobile devices, electronic health policy, privacy, ethical socio-economic inequalities, and legal risks. In addition, there is a lack of supporting ICT infrastructure, WIFI, GPS services, and abuse of contact tracing apps (Mbunge, 2020). Recently, the effective use of artificial intelligence solutions in the medical area has been the black-box models because medical professionals do not fully understand the logic of a particular machine prediction. Utilizing multi-class disease segmentation and in-depth analysis of the characteristics of each class and their association with severity. AI can assist the community in various ways, including early warnings and alerts, diagnosis and prognosis, tracking and prediction, treatments and cures, data dashboards, and social control, by prioritizing individuals for testing and thus increasing the rate at which positive individuals can be identified. Ilyas et al. (2020)

## Conclusion

In conclusion, although it is very hard to find any positive impact of the COVID-19 pandemic on most of the sectors that touched our lives, from sociological and health perspectives to the economic crash, and at personal and community levels, one can appraise the huge effort made by the scientific community in an attempt to alleviate such disastrous impact. This survey covered the main technical contributions from data mining perspectives, focusing on social data, contact tracing, medical imaging, and health-related time-series data. We presented the challenges, techniques, and open problems with opportunities that can be tackled soon. For instance, social data mining needs deeper correlation and semantic analysis with other data types, such as health and contact tracing data. Contact tracing, on the other hand, could not be widely adopted because of the large privacy concerns and effectiveness of current solutions. Finally, research on medical imaging has achieved great support for the automatic early screening of infected cases, but deeper pattern recognition and tracking of the disease in order to predict the best treatment ahead of time can immensely enrich the current solutions. Overall, the COVID-19 pandemic has highlighted the importance of data mining techniques in analyzing large volumes of data in real-time, integrating data from multiple sources, developing predictive models, ensuring data quality, and considering ethical considerations. This survey is unique based on all these perspectives and recommendations and can leverage more advancements in the related fields.

**Author Contributions** All authors contribute equally.

**Funding** Not Applicable.

**Data Availability Statement** Data is available from the authors upon reasonable request.

## Declarations

**Conflict of Interest** The authors declare that there is no conflict of interest regarding the publication of this paper.

**Ethical Approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**Informed Consent** Informed consent was obtained from all individual participants included in the study.

## References

- Abdalla, W., Renukappa, S., & Suresh, S. (2023). Managing covid-19-related knowledge: A smart cities perspective. *Knowledge and Process Management*, 30(1), 87–109.
- Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the covid-19 pandemic: Infoveillance study. *Journal of Medical Internet Research*, 22(4), e19016.
- Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the covid-19 pandemic: Infoveillance study. *Journal of Medical Internet Research*, 22(4), e19016.
- Abdul-Mageed, M., & Diab, M. T. (2011). Subjectivity and sentiment annotation of modern standard arabic newswire. In: *Proceedings of the 5th linguistic annotation workshop*, pp. 110–118.
- Abdul-Mageed, M., & Diab, M., (2014) SANA: A large scale multi-genre, multi-dialect lexicon for Arabic subjectivity and sentiment analysis. In: *Proceedings of the ninth international conference on Language Resources and Evaluation (LREC'14)*, European Language Resources Association (ELRA), Reykjavik, Iceland, pp. 1162–1169.
- Abuhammad, S., Khabour, O. F., & Alzoubi, K. H. (2020). Covid-19 contact-tracing technology: Acceptability and ethical issues of use. *Patient Preference and Adherence*, 14, 1639.
- Adly, A. S., Adly, A. S., & Adly, M. S. (2020). Approaches based on artificial intelligence and the internet of intelligent things to prevent the spread of covid-19: Scoping review. *Journal of Medical Internet Research*, 22(8), e19104.
- Agarwal, A., Salehundam, P., Padhee, S., Romine, W. L., & Banerjee, T. (2020). Leveraging natural language processing to mine issues on twitter during the covid-19 pandemic. [arXiv:2011.00377](https://arxiv.org/abs/2011.00377)
- Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S., Seneviratne, A., Hu, W., Janicke, H., & Jha, S. K. (2020). A survey of covid-19 contact tracing apps. *IEEE Access*, 8, 134577–134601.
- Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S., Seneviratne, A., Hu, W., Janicke, H., & Jha, S. K. (2020). A survey of covid-19 contact tracing apps. *IEEE Access*, 8, 134577–134601.
- Ajaz, F., Naseem, M., Sharma, S., Shabaz, M., & Dhiman, G. (2022). Covid-19: Challenges and its technological solutions using iot. *Current Medical Imaging*, 18(2), 113–123.
- Alamoodi, A., Zaidan, B., Zaidan, A., Albahri, O., Mohammed, K., Malik, R., Almahdi, E., Chyad, M., Tareq, Z., Albahri, A., et al. (2020). Sentiment analysis and its applications in fighting covid-19 and infectious diseases: A systematic review. *Expert Systems with Applications*, 114155.
- Alanazi, E., Alashaikh, A., Alqurashi, S., & Alanazi, A. (2020). Identifying and ranking common covid-19 symptoms from tweets in Arabic: Content analysis. *Journal of Medical Internet Research*, 22(11), e21329.
- Alarabi, L., Basalamah, S., Hendawi, A., Abdalla, M. (2021). Traceall: A real-time processing for contact tracing using indoor trajectories. *Information*, 12(5). <https://doi.org/10.3390/info12050202>, <https://www.mdpi.com/2078-2489/12/5/202>

- Allelyani, M., Alghamdi, A., Shubayr, N., Alashban, Y., Almater, H., Alamri, S., & Alghamdi, A. J. (2021). The impact of the covid-19 pandemic on medical imaging case volumes in aseer region: A retrospective study. *Medicines*, *8*(11), 70.
- Alqurashi, S., Alhindi, A., & Alanazi, E. (2020). Large arabic twitter dataset on covid-19. [arXiv:2004.04315](https://arxiv.org/abs/2004.04315)
- Alqurashi, S., Alhindi, A., & Alanazi, E. (2020). Large arabic twitter dataset on covid-19. [arXiv:2004.04315](https://arxiv.org/abs/2004.04315)
- Al-Rawi, A., & Shukla, V. (2020). Bots as active news promoters: A digital analysis of covid-19 tweets. *Information*, *11*(10), 461.
- Al-Rawi, A., & Shukla, V. (2020). Bots as active news promoters: A digital analysis of covid-19 tweets. *Information*, *11*(10), 461.
- Alsudias, L., & Rayson, P. (2020). Covid-19 and arabic twitter: How can arab world governments and public health organizations learn from social media?. In: *Proceedings of the 1st workshop on NLP for COVID-19 at ACL 2020*.
- Alsudias, L., & Rayson, P. (2020). COVID-19 and Arabic Twitter: How can Arab world governments and public health organizations learn from social media? In: *Proceedings of the 1st workshop on NLP for COVID-19 at ACL 2020*, Association for Computational Linguistics, Online. <https://www.aclweb.org/anthology/2020.nlpCOVID19-acl.16>
- Alzahrani, S. I., Aljamaan, I. A., & Al-Fakih, E. A. (2020) Forecasting the spread of them covid-19 pandemic in Saudi Arabia using arima prediction model under current public health interventions. *Journal of Infection and Public Health*, *13*(7) 914–919.
- Alzahrani, S. I., Aljamaan, I. A., & Al-Fakih, E. A. (2020). Forecasting the spread of the covid-19 pandemic in Saudi Arabia using arima prediction model under current public health interventions. *Journal of Infection and Public Health*, *13*(7), 914–919.
- Amram, O., Amiri, S., Lutz, R. B., Rajan, B., & Monsivais, P. (2020). Development of a vulnerability index for diagnosis with the novel coronavirus, covid-19, in Washington State, USA. *Health & Place*.
- Anastassopoulou, C., Russo, L., Tsakris, A., & Siettos, C. (2020). Data-based analysis, modelling and forecasting of the covid-19 outbreak. *PLoS one*, *15*(3), e0230405.
- Annas, S., Pratama, M. I., Rifandi, M., Sanusi, W., & Side, S. (2020). Stability analysis and numerical simulation of seir model for pandemic covid-19 spread in Indonesia. *Chaos, Solitons & Fractals*, *139*, 110072.
- Anshari, M., Hamdan, M., Ahmad, N., Ali, E., & Haidi, H. (2023). Covid-19, artificial intelligence, ethical challenges and policy implications. *Ai & Society*, *38*(2), 707–720.
- Apuke, O. D., & Omar, B. (2021). Fake news and covid-19: Modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, *56*, 101475.
- Arunmozhi, M., Persis, J., Sreedharan, V. R., Chakraborty, A., Zouadi, T., & Khamlichi, H. (2022). Managing the resource allocation for the covid-19 pandemic in healthcare institutions: A pluralistic perspective. *International Journal of Quality & Reliability Management*, *39*(9), 2184–2204.
- Ayoub, J., Yang, X. J., & Zhou, F. (2021). Combat covid-19 infodemic using explainable natural language processing models. *Information Processing & Management*, *58*(4), 102569. <https://doi.org/10.1016/j.ipm.2021.102569>, <https://www.sciencedirect.com/science/article/pii/S0306457321000704>
- Aytaç, U. C., Güneş, A., & Ajlouni, N. (2022). A novel adaptive momentum method for medical image classification using convolutional neural network. *BMC Medical Imaging*, *22*(1), 1–12.
- Bahja, M., Hammad, R., Kuhail, M. A. (2020). Capturing public concerns about coronavirus using arabic tweets: An nlp-driven approach. In: *2020 IEEE/ACM 13th International Conference on Utility and Cloud Computing (UCC)*, *IEEE*, pp. 310–315.
- Bayham, J., & Fenichel, E. P. (2020). The impact of school closure for covid-19 on the us healthcare workforce and the net mortality effects. *Available at SSRN* 3555259.
- Beare, B. K., & Toda, A. A. (2020). On the emergence of a power law in the distribution of covid-19 cases. *Physica D: Nonlinear Phenomena*, *412*, 132649.
- Bentotahewa, V., Hewage, C., & Williams, J. (2021). Solutions to big data privacy and security challenges associated with covid-19 surveillance systems. *Frontiers in Big Data*, *4*, 645204.
- Bhattacharjee, S. (2020). Statistical investigation of relationship between spread of coronavirus disease (covid-19) and environmental factors based on study of four mostly affected places of China and five mostly affected places of Italy. [arXiv:2003.11277](https://arxiv.org/abs/2003.11277)
- Bhattacharya, S., Maddikunta, P. K. R., Pham, Q.-V., Gadekallu, T. R., Chowdhary, C. L., Alazab, M., Piran, M. J., et al. (2021). Deep learning and medical image processing for coronavirus (covid-19) pandemic: A survey. *Sustainable Cities and Society*, *65*, 102589.

- Born, J., Beymer, D., Rajan, D., Coy, A., Mukherjee, V. V., Manica, M., Prasanna, P., Ballah, D., Guindy, M., Shaham, D. et al. (2021). On the role of artificial intelligence in medical imaging of covid-19. *Patterns*, 2(6).
- Boyle, F., & Sherman, D. (2006). Scopus™: The product and its development. *The Serials Librarian*, 49(3), 147–153.
- Bradshaw, W. J., Alley, E. C., Huggins, J. H., Lloyd, A. L., & Esvelt, K. M. (2021). Bidirectional contact tracing could dramatically improve covid-19 control. *Nature Communications*, 12(1), 1–9.
- Braithwaite, I., Callender, T., Bullock, M., & Aldridge, R. W. (2020). Automated and partly automated contact tracing: A systematic review to inform the control of covid-19. *The Lancet Digital Health*, 2(11).
- Capasso, A., Kim, S., Ali, S. H., Jones, A. M., DiClemente, R. J., & Tozan, Y. (2022). Employment conditions as barriers to the adoption of covid-19 mitigation measures: How the covid-19 pandemic may be deepening health disparities among low-income earners and essential workers in the united states. *BMC Public Health*, 22(1), 1–13.
- Castex, G., Dechter, E., & Lorca, M. (2020). Covid-19: The impact of social distancing policies, cross-country analysis. *Economics of Disasters and Climate Change*, 1–25.
- Castro, M. C., de Carvalho, L. R., Chin, T., Kahn, R., Franca, G. V., Macario, E. M., & de Oliveira, W. K. (2020). Demand for hospitalization services for covid-19 patients in Brazil. *MedRxiv*.
- Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment analysis of covid-19 tweets by deep learning classifiers-a study to show how popularity is affecting accuracy in social media. *Applied Soft Computing*, 97, 106754.
- Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment analysis of covid-19 tweets by deep learning classifiers-a study to show how popularity is affecting accuracy in social media. *Applied Soft Computing*, 97, 106754.
- Chan, E. Y., & Saqib, N. U. (2021). Privacy concerns can explain unwillingness to download and use contact tracing apps when covid-19 concerns are high. *Computers in Human Behavior*, 119, 106718.
- Chao, H., Fang, X., Zhang, J., Homayounieh, F., Arru, C. D., Digumarthy, S. R., Babaei, R., Mobin, H. K., Mohseni, I., Saba, L., et al. (2021). Integrative analysis for covid-19 patient outcome prediction. *Medical Image Analysis*, 67, 101844.
- Chen, T., Rong, J., Peng, L., Yang, J., Cong, G., Fang, J. (2021). Analysis of social effects on employment promotion policies for college graduates based on data mining for online use review in china during the covid-19 pandemic. In: *Healthcare*, Multidisciplinary Digital Publishing Institute, 9, p. 846.
- Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance*, 6(2), e19273.
- Chernozhukov, V., Kasahara, H., & Schrimpf, P. (2021). Causal impact of masks, policies, behavior on early covid-19 pandemic in the US. *Journal of Econometrics*, 220(1), 23–62.
- Chierigato, M., Frangiamore, F., Morassi, M., Baresi, C., Nici, S., Bassetti, C., Bnà, C., & Galelli, M. (2022). A hybrid machine learning/deep learning covid-19 severity predictive model from ct images and clinical data. *Scientific Reports*, 12(1), 1–15.
- Chiroma, H., Ezugwu, A. E., Jauro, F., Al-Garadi, M. A., Abdullahi, I. N., & Shuib, L. (2020). Early survey with bibliometric analysis on machine learning approaches in controlling covid-19 outbreaks. *PeerJ Computer Science*, 6, e313.
- Cho, H., Ippolito, D., & Yu, Y. W. (2020). Contact tracing mobile apps for covid-19: Privacy considerations and related trade-offs. [arXiv:2003.11511](https://arxiv.org/abs/2003.11511)
- Chowdhury, N. K., Rahman, M. M., & Kabir, M. A. (2020). Pdcovidnet: A parallel-dilated convolutional neural network architecture for detecting covid-19 from chest x-ray images. *Health Information Science and Systems*, 8(1), 1–14.
- Cinelli, M., Quattrocioni, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The covid-19 social media infodemic. *Scientific Reports*, 10(1), 1–10.
- Colizza, V., Grill, E., Mikolajczyk, R., Cattuto, C., Kucharski, A., Riley, S., Kendall, M., Lythgoe, K., Bon-sall, D., Wymant, C., et al. (2021). Time to evaluate covid-19 contact-tracing apps. *Nature Medicine*, 27(3), 361–362.
- Connor, C., De Valliere, N., Warwick, J., Stewart-Brown, S., & Thompson, A. (2022). The cov-ed survey: Exploring the impact of learning and teaching from home on parent/carers' and teachers' mental health and wellbeing during covid-19 lockdown. *BMC Public Health*, 22(1), 1–15.

- Cortés-Martínez, K. V., Estrada-Esquivel, H., Martínez-Rebollar, A., Hernández-Pérez, Y., & Ortiz-Hernández, J. (2022). The state of the art of data mining algorithms for predicting the covid-19 pandemic. *Axioms*, *11*(5), 242.
- COVID, T. I., Reiner, R., Barber, R., & Collins, J. (2020). Modeling covid-19 scenarios for the United States. *Nature medicine*.
- Cuan-Baltazar, J. Y., Muñoz-Perez, M. J., Robledo-Vega, C., Pérez-Zepeda, M. F., & Soto-Vega, E. (2020). Misinformation of covid-19 on the internet: Infodemiology study. *JMIR Public Health and Surveillance*, *6*(2), e18444.
- Cuello-García, C., Pérez-Gaxiola, G., & van Amelsvoort, L. (2020). Social media can have an impact on how we manage and investigate the covid-19 pandemic. *Journal of Clinical Epidemiology*, *127*, 198–201.
- Dar, A. B., Lone, A. H., Zahoor, S., Khan, A. A., & Naaz, R. (2020). Applicability of mobile contact tracing in fighting pandemic (covid-19): Issues, challenges and solutions. *Computer Science Review*, *38*, 100307. <https://doi.org/10.1016/j.cosrev.2020.100307>, [www.sciencedirect.com/science/article/pii/S157401372030407X](http://www.sciencedirect.com/science/article/pii/S157401372030407X)
- Dash, S., Chakraborty, C., Giri, S. K., & Pani, S. K. (2021). Intelligent computing on time-series data analysis and prediction of covid-19 pandemics. *Pattern Recognition Letters*, *151*, 69–75.
- de Figueiredo, C. S., Sandre, P. C., Portugal, L. C. L., Mázala-de Oliveira, T., da Silva Chagas, L., Raony, Í., Ferreira, E. S., Giestal-de Araujo, E., Dos Santos, A. A., & Bomfim, P.O.-S. (2021). Covid-19 pandemic impact on children and adolescents' mental health: Biological, environmental, and social factors. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *106*, 110171.
- De Santis, E., Martino, A., & Rizzi, A. (2020). An infoveillance system for detecting and tracking relevant topics from Italian tweets during the covid-19 event. *IEEE Access*, *8*, 132527–132538.
- Desai, P. S. (2021). News sentiment informed time-series analyzing ai (sitala) to curb the spread of covid-19 in Houston. *Expert Systems with Applications*, *180*, 115104. <https://doi.org/10.1016/j.eswa.2021.115104>, [www.sciencedirect.com/science/article/pii/S0957417421005455](http://www.sciencedirect.com/science/article/pii/S0957417421005455)
- Devi, V. A., & Nayyar, A. (2021). Evaluation of geotagging twitter data using sentiment analysis during covid-19. In: *Proceedings of the second international conference on information management and machine intelligence*, Springer, pp. 601–608.
- Devi, V. A., & Nayyar, A. (2021). Evaluation of geotagging twitter data using sentiment analysis during covid-19. In: *Proceedings of the second international conference on information management and machine intelligence*, Springer, pp. 601–608.
- Dimitrov, D., Baran, E., Fafalios, P., Yu, R., Zhu, X., Zloch, M., & Dietze, S. (2020). Tweetscov19-a knowledge base of semantically annotated tweets about the covid-19 pandemic. In: *Proceedings of the 29th ACM international conference on information & knowledge management*, pp. 2991–2998.
- Dimitrov, D., Baran, E., Fafalios, P., Yu, R., Zhu, X., Zloch, M., & Dietze, S. (2020). Tweetscov19-a knowledge base of semantically annotated tweets about the covid-19 pandemic. In: *Proceedings of the 29th ACM international conference on information & knowledge management*, pp. 2991–2998.
- Durowaye, T. D., Rice, A. R., Konkle, A., & Phillips, K. P. (2022). Public health perinatal promotion during covid-19 pandemic: A social media analysis. *BMC Public Health*, *22*(1), 1–12.
- Elnagar, A., Al-Debsi, R., & Einea, O. (2020). Arabic text classification using deep learning models. *Information Processing & Management*, *57*(1), 102121.
- Elsheikh, A. H., Saba, A. I., Abd Elaziz, M., Lu, S., Shanmugan, S., Muthuramalingam, T., Kumar, R., Mosleh, A. O., Essa, F., & Shehabeldeen, T. A. (2021). Deep learning-based forecasting model for covid-19 outbreak in Saudi Arabia. *Process Safety and Environmental Protection*, *149*, 223–233.
- Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., et al. (2020). Impact of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand. imperial college covid-19 response team. *Imperial College COVID-19 Response Team*, *20*.
- Gao, J., Zheng, P., Jia, Y., Chen, H., Mao, Y., Chen, S., Wang, Y., Fu, H., & Dai, J. (2020). Mental health problems and social media exposure during covid-19 outbreak. *Plos one*, *15*(4), e0231924.
- Gencoglu, O. (2020). Large-scale, language-agnostic discourse classification of tweets during covid-19. *Machine Learning and Knowledge Extraction*, *2*(4), 603–616.
- Gencoglu, O. (2020). Large-scale, language-agnostic discourse classification of tweets during covid-19. *Machine Learning and Knowledge Extraction*, *2*(4), 603–616.
- Ghosh, S., & Das, L. C. (2022). Using data mining techniques for covid-19: A systematic. *Science and Technology*, *8*(2), 36–42.

- Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., Colaneri, M. (2020). Modelling the covid-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine* 26(6), 855–860.
- Gozes, O., Frid-Adar, M., Greenspan, H., Browning, P. D., Zhang, H., Ji, W., Bernheim, A., & Siegel, E. (2020). Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. [arXiv:2003.05037](https://arxiv.org/abs/2003.05037)
- Grasselli, G., Pesenti, A., & Cecconi, M. (2020). Critical care utilization for the covid-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. *Jama*, 323(16), 1545–1546.
- Guntuku, S. C., Sherman, G., Stokes, D. C., Agarwal, A. K., Seltzer, E., Merchant, R. M., & Ungar, L. H. (2020). Tracking mental health and symptom mentions on twitter during covid-19. *Journal of General Internal Medicine*, 35(9), 2798–2800.
- Gupta, R., Ibraheim, M. K., & Doan, H. Q. (2020). Teledermatology in the wake of covid-19: Advantages and challenges to continued care in a time of disarray. *Journal of the American Academy of Dermatology*, 83(1), 168–169.
- Hamzah, F. B., Lau, C., Nazri, H., Ligot, D., Lee, G., Tan, C., Shaib, M., Zaidon, U., Abdullah, A., Chung, M., et al. (2020). Coronatracker: Worldwide covid-19 outbreak data analysis and prediction. *Bull World Health Organ*, 1(32).
- Haouari, F., Hasanain, M., Suwaileh, R., & Elsayed, T. (2021). ArCOV-19: The first Arabic COVID-19 Twitter dataset with propagation networks. In: *Proceedings of the sixth arabic natural language processing workshop, association for computational linguistics*, pp. 82–91.
- Heikal, M., Torki, M., & El-Makky, N. (2018). Sentiment analysis of arabic tweets using deep learning. *Procedia Computer Science*, 142, 114–122.
- Hernandez-Matamoros, A., Fujita, H., Hayashi, T., & Perez-Meana, H. (2020). Forecasting of covid19 per regions using arima models and polynomial functions. *Applied Soft Computing*, 96, 106610–106610.
- Ho, K. K., Chiu, D. K., & Sayama, K. C. (2023). When privacy, distrust, and misinformation cause worry about using covid-19 contact-tracing apps. *IEEE Internet Computing*, 01, 1–7.
- Hossain, M., Junus, A., Zhu, X., Jia, P., Wen, T. -H., Pfeiffer, D., & Yuan, H. -Y. (2020). The effects of border control and quarantine measures on global spread of covid-19, Alvin and Zhu, Xiaolin and Jia, Pengfei and Wen, Tzai-Hung and Pfeiffer, Dirk and Yuan, Hsiang-Yu. *The Effects of Border Control and Quarantine Measures on Global Spread of COVID-19* (3/6/2020).
- Hou, K., Hou, T., & Cai, L. (2021). Public attention about covid-19 on social media: An investigation based on data mining and text analysis. *Personality and Individual Differences*, 175, 110701.
- Hussain, A., & Sheikh, A. (2021). Opportunities for artificial intelligence-enabled social media analysis of public attitudes toward covid-19 vaccines. *NEJM Catalyst Innovations in Care Delivery*, 2(1).
- Ibrahim, H. S., Abdou, S. M., & Gheith, M. (2015). Sentiment analysis for modern standard arabic and colloquial. [arXiv:1505](https://arxiv.org/abs/1505)
- Ilyas, M., Rehman, H., & Nait-Ali, A. (2020). Detection of covid-19 from chest x-ray images using artificial intelligence: An early review. [arXiv:2004.05436](https://arxiv.org/abs/2004.05436)
- Iwendi, C., Mohan, S., Ibeke, E., Ahmadian, A., Ciano, T., et al. (2022). Covid-19 fake news sentiment analysis. *Computers and Electrical Engineering*, 101, 107967.
- Jain, R., Gupta, M., Taneja, S., & Hemanth, D. J. (2020). Deep learning based detection and analysis of covid-19 on chest x-ray images. *Applied Intelligence*, 1–11.
- Jamieson, J., Yamashita, N., Epstein, D. A., & Chen, Y. (2021). Deciding if and how to use a covid-19 contact tracing app: Influences of social factors on individual use in Japan. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–30.
- Janarthanan, S., Rajendran, M., Biju, T. S., Ravi, N., Sundaramoorthy, K., & Nandan Mohanty, S. (2021). Artificial intelligence (ai) combined with medical imaging enables rapid diagnosis for covid-19. In: *Applications of artificial intelligence in COVID-19*, Springer, pp. 55–72.
- Kabir, M. Y., & Madria, S. (2021). Emocov: Machine learning for emotion detection, analysis and visualization using covid-19 tweets. *Online Social Networks and Media*, 23, 100135. <https://doi.org/10.1016/j.osnem.2021.100135>, <https://www.sciencedirect.com/science/article/pii/S2468696421000197>
- Kang, E., Lee, S. Y., Jung, H., Kim, M. S., Cho, B., & Kim, Y. S. (2020). Operating protocols of a community treatment center for isolation of patients with coronavirus disease, South Korea. *Emerging Infectious Diseases*, 26(10), 2329.
- Katris, C. (2021). A time series-based statistical approach for outbreak spread forecasting: Application of covid-19 in Greece. *Expert Systems with Applications*, 166, 114077.

- Kiamari, M., Ramachandran, G., Nguyen, Q., Pereira, E., Holm, J., & Krishnamachari, B. (2020). Covid-19 risk estimation using a time-varying sir-model *In: Proceedings of the 1st ACM SIGSPATIAL international workshop on modeling and understanding the spread of COVID-19*, pp. 36–42.
- Kim, K.-M., & Rhee, H.-S. (2022). Influential factors for covid-19 related distancing in daily life: A distinct focus on ego-gram. *BMC Public Health*, 22(1), 1–13.
- Koh, J. X., & Liew, T. M. (2020). How loneliness is talked about in social media during covid-19 pandemic: Text mining of 4,492 twitter feeds. *Journal of Psychiatric Research*. <https://doi.org/10.1016/j.jpsychires.2020.11.015>, [www.sciencedirect.com/science/article/pii/S0022395620310748](http://www.sciencedirect.com/science/article/pii/S0022395620310748)
- Koh, J. X., & Liew, T. M. (2020). How loneliness is talked about in social media during covid-19 pandemic: Text mining of 4,492 twitter feeds. *Journal of Psychiatric Research*.
- Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., Sun, F., Jit, M., Munday, J. D., et al. (2020). Early dynamics of transmission and control of covid-19: A mathematical modelling study. *The Lancet Infectious Diseases*, 20(5), 553–558.
- Kuo, C.-P., & Fu, J. S. (2021). Evaluating the impact of mobility on covid-19 pandemic with machine learning hybrid predictions. *Science of The Total Environment*, 758, 144151.
- Lai, S., Bogoch, I. I., Ruktanonchai, N. W., Watts, A., Lu, X., Yang, W., Yu, H., Khan, K., & Tatem, A. J. (2020). Assessing spread risk of wuhan novel coronavirus within and beyond China, January-April : A travel network-based modelling study, *MedRxiv*.
- Lamsal, R. (2020). Coronavirus (covid-19) geo-tagged tweets dataset. <https://doi.org/10.21227/fpsb-jz61>
- Lamsal, R. (2020). Coronavirus (covid-19) tweets dataset. <https://doi.org/10.21227/781w-ef42>
- Lamsal, R. (2020). Design and analysis of a large-scale covid-19 tweets dataset. *Applied Intelligence*, 1–15.
- Lazarus, J. V., Ratzan, S. C., Palayew, A., Gostin, L. O., Larson, H. J., Rabin, K., Kimball, S., & El-Mohandes, A. (2021). A global survey of potential acceptance of a covid-19 vaccine. *Nature Medicine*, 27(2), 225–228.
- Lee, H. S. (2020). Exploring the initial impact of covid-19 sentiment on us stock market using big data. *Sustainability*, 12(16), 6648.
- Leung, C. K., Kaufmann, T. N., Wen, Y., Zhao, C., & Zheng, H. (2022). Revealing covid-19 data by data mining and visualization, in: *Advances in Intelligent Networking and Collaborative Systems: The 13th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2021)*, Springer, 13 pp. 70–83.
- Leung, K., Wu, J. T., Liu, D., & Leung, G. M. (2020). First-wave covid-19 transmissibility and severity in China outside hubei after control measures, and second-wave scenario planning: A modelling impact assessment. *The Lancet*, 395(10233), 1382–1393.
- Li, L., Yang, Z., Dang, Z., Meng, C., Huang, J., Meng, H., Wang, D., Chen, G., Zhang, J., Peng, H., et al. (2020). Propagation analysis and prediction of the covid-19. *Infectious Disease Modelling*, 5, 282–292.
- Li, C., Chen, L. J., Chen, X., Zhang, M., Pang, C. P., & Chen, H. (2020). Retrospective analysis of the possibility of predicting the covid-19 outbreak from internet searches and social media data, China, 2020. *Eurosurveillance*, 25(10), 2000199.
- Liang, W., Fan, Y., Li, K.-C., Zhang, D., & Gaudiot, J.-L. (2020). Secure data storage and recovery in industrial blockchain network environments. *IEEE Transactions on Industrial Informatics*, 16(10), 6543–6552.
- Lin, L., & Hou, Z. (2020). Combat covid-19 with artificial intelligence and big data. *Journal of Travel Medicine*, 27(5), taaa080.
- Liu, P., Beeler, P., & Chakrabarty, R. K. (2020). Covid-19 progression timeline and effectiveness of response-to-spread interventions across the united states, *medRxiv*.
- Liu, M., Zhang, Z., Chai, W., & Wang, B. (2023). Privacy-preserving covid-19 contact tracing solution based on blockchain. *Computer Standards & Interfaces*, 83, 103643.
- López, V., & Čukić, M. (2021). A dynamical model of sars-cov-2 based on people flow networks. *Safety Science*, 134, 105034.
- Lucivero, F., Hallowell, N., Johnson, S., Prainsack, B., Samuel, G., & Sharon, T. (2020). Covid-19 and contact tracing apps: Ethical challenges for a social experiment on a global scale. *Journal of Bioethical Inquiry*, 17(4), 835–839.
- Luo, Y., Li, W., Zhao, T., Yu, X., Zhang, L., Li, G., & Tang, N. (2020). Deeptrack: Monitoring and exploring spatio-temporal data: A case of tracking covid-19. *Proceedings of the VLDB Endowment*, 13(12), 2841–2844.



- Luz, E., Silva, P., Silva, R., Silva, L., Guimarães, J., Miozzo, G., Moreira, G., & Menotti, D. (2021). Towards an effective and efficient deep learning model for covid-19 patterns detection in x-ray images. *Research on Biomedical Engineering*, 1–14.
- Mahalle, P., Kalamkar, A. B., Dey, N., Chaki, J., Shinde, G. R., et al. (2020). Forecasting models for coronavirus (covid-19): A survey of the state-of-the-art.
- Mahmud, T., Rahman, M. A., & Fattah, S. A. (2020). Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization. *Computers in Biology and Medicine*, 122, 103869.
- Mavragani, A. (2020). Tracking covid-19 in europe: Infodemiology approach. *JMIR Public Health and Surveillance*, 6(2), e18941.
- Mbunge, E. (2020). Integrating emerging technologies into covid-19 contact tracing: Opportunities, challenges and pitfalls. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(6), 1631–1636.
- Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., & Soufi, G. J. (2020). Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning. *Medical Image Analysis*, 65, 101794.
- Moghadas, S. M., Shoukat, A., Fitzpatrick, M. C., Wells, C. R., Sah, P., Pandey, A., Sachs, J. D., Wang, Z., Meyers, L. A., Singer, B. H. (2020) et al. Projecting hospital utilization during the covid-19 outbreaks in the United States. *Proceedings of the National Academy of Sciences*, 117(16) 9122–9126.
- Mokbel, M., Abbar, S., & Stanojevic, R. (2020). Contact tracing: Beyond the apps. *SIGSPATIAL Special*, 12(2), 15–24.
- Mourad, A., & Darwish, K. (2013). Subjectivity and sentiment analysis of modern standard arabic and arabic microblogs. In: *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pp. 55–64.
- Murphy, R., Calugi, S., Cooper, Z., & Dalle Grave, R. (2020). Challenges and opportunities for enhanced cognitive behaviour therapy (cbt-e) in light of covid-19. *The Cognitive Behaviour Therapist*, 13.
- Mushtaq, M. F., Fareed, M. M. S., Almutairi, M., Ullah, S., Ahmed, G., & Munir, K. (2022). Analyses of public attention and sentiments towards different covid-19 vaccines using data mining techniques. *Vaccines*, 10(5), 661.
- Mutlu, E. C., Oghaz, T., Jasser, J., Tutunculer, E., Rajabi, A., Tayebi, A., Ozmen, O., & Garibay, I. (2020). A stance data set on polarized conversations on twitter about the efficacy of hydroxychloroquine as a treatment for covid-19. *Data in brief*, 33, 106401.
- Mutlu, E., Oghaz, T., Jasser, J., Tutunculer, E., Rajabi, A., Tayebi, A., Ozmen, O., & Garibay, I. (2020). A stance data set on polarized conversations on twitter about the efficacy of hydroxychloroquine as a treatment for covid-19. *Data in Brief*, 33, 106401–106401.
- Nadim, S. S., Ghosh, I., & Chattopadhyay, J. (2021). Short-term predictions and prevention strategies for covid-19: a model-based study. *Applied Mathematics and Computation*, 404, 126251.
- Nakov, P., & Da San Martino, G. (2021). Fake news, disinformation, propaganda, media bias, and flattening the curve of the covid-19 infodemic. In: *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 4054–4055.
- Namasudra, S., Dhamodharavadhani, S., & Rathipriya, R. (2023). Nonlinear neural network based forecasting model for predicting covid-19 cases. *Neural Processing Letters*, 1–21.
- Naseem, U., Razzak, I., Khushi, M., Eklund, P. W., & Kim, J. (2021). CoviSenti: A large-scale benchmark twitter data set for covid-19 sentiment analysis. *IEEE Transactions on Computational Social Systems*.
- Nemes, L., & Kiss, A. (2021). Social media sentiment analysis based on covid-19. *Journal of Information and Telecommunication*, 5(1), 1–15.
- Oehmke, T. B., Post, L. A., Moss, C. B., Issa, T. Z., Boctor, M. J., Welch, S. B., & Oehmke, J. F. (2021). Dynamic panel data modeling and surveillance of covid-19 in metropolitan areas in the united states: Longitudinal trend analysis. *Journal of Medical Internet Research*, 23(2), e26081.
- Oliveira, J. F., Jorge, D. C., Veiga, R. V., Rodrigues, M. S., Torquato, M. F., da Silva, N. B., Fiaccone, R. L., Cardim, L. L., Pereira, F. A., de Castro, C. P. et al. (2021). Mathematical modeling of covid-19 in 14.8 million individuals in Bahia, Brazil. *Nature Communications* 12(1), 1–13.
- Ordun, C., Purushotham, S., & Raff, E. (2020). Exploratory analysis of covid-19 tweets using topic modeling, umap, and digraphs. [arXiv:2005.03082](https://arxiv.org/abs/2005.03082)
- Organization, W. H., et al. (2021). Looking back at a year that changed the world: Who's response to covid-19, 22 January 2021. *Tech. rep.: World Health Organization*.
- Ouchicha, C., Ammor, O., & Meknassi, M. (2020). Cvdnet: A novel deep learning architecture for detection of coronavirus (covid-19) from chest x-ray images. *Chaos, Solitons & Fractals*, 140, 110245–110245.

- Padhan, R., & Prabheesh, K. (2021). The economics of covid-19 pandemic: A survey. *Economic Analysis and Policy*, 70, 220–237.
- Park, Y. J., Choe, Y. J., Park, O., Park, Kim, S.Y., Kim, J., Kweon, S., Woo, Y., Gwack, J., Kim, S. S., et al. (2020). 1440 Contact tracing during coronavirus disease outbreak, South Korea, 2020. *Emerging Infectious Diseases*, 26(10), 2465–2468.
- Park, J. Y., Mistur, E., Kim, D., Mo, Y., Hoefler, R. (2021). Toward human-centric urban infrastructure: Text mining for social media data to identify the public perception of covid-19 policy in transportation hubs. *Sustainable Cities and Society*, 103524.
- Park, Y.-E. (2022). Developing a covid-19 crisis management strategy using news media and social media in big data analytics. *Social Science Computer Review*, 40(6), 1358–1375.
- Perumal, V., Narayanan, V., & Rajasekar, S. J. S. (2020). Detection of covid-19 using cxr and ct images using transfer learning and haralick features. *Applied Intelligence*, 1–18.
- Pham, D. P. T., Quang, A. H. N., & Duong, D. (2022). The impact of us presidents on market returns: Evidence from trump's tweets. *Research in International Business and Finance*, 101681.
- Pirkis, J., John, A., Shin, S., DelPozo-Banos, M., Arya, V., Analuisa-Aguilar, P., Appleby, L., Arensman, E., Bantjes, J., Baran, A., et al. (2021). Suicide trends in the early months of the covid-19 pandemic: An interrupted time-series analysis of preliminary data from 21 countries. *The Lancet Psychiatry*, 8(7), 579–588.
- Qazi, U., Imran, M., & Ofli, F. (2020). Geocov19: A dataset of hundreds of millions of multilingual covid-19 tweets with location information. *SIGSPATIAL Special*, 12(1), 6–15.
- Qazi, U., Imran, M., & Ofli, F. (2020). Geocov19: A dataset of hundreds of millions of multilingual covid-19 tweets with location information. *SIGSPATIAL Special*, 12(1), 6–15.
- Quak, E., Girault, G., Thenint, M. A., Weyts, K., Lequesne, J., & Lasnon, C. (2021). Author gender inequality in medical imaging journals and the covid-19 pandemic. *Radiology* 204417.
- Rehouma, R., Buchert, M., & Chen, Y.-P. P. (2021). Machine learning for medical imaging-based covid-19 detection and diagnosis. *International Journal of Intelligent Systems*, 5085–5115.
- Rocha Filho, T. M., dos Santos, F. S. G., Gomes, V. B., Rocha, T. A., Croda, J. H., Ramalho, W. M., Araujo, W. N. (2020). Expected impact of covid-19 outbreak in a major metropolitan area in Brazil. *MedRxiv*.
- Rovetta, A., & Bhagavathula, A. S. (2020). Covid-19-related web search behaviors and infodemic attitudes in italy: Infodemiological study. *JMIR Public Health and Surveillance*, 6(2), e19374.
- Russo, L., Anastassopoulou, C., Tsakris, A., Bifulco, G., Campana, E., Toraldo, G., Siettos, C., (2020). T. DAY-ZERO, forecasting the fade out of the covid-19 outbreak in lombardy, Italy: A compartmental modelling and numerical optimization approach. *MedRxiv*.
- Sadler, T. D., Friedrichsen, P., Zangori, L., & Ke, L. (2020). Technology-supported professional development for collaborative design of covid-19 instructional materials. *Journal of Technology and Teacher Education*, 28(2), 171–177.
- Safdari, R., Rezayi, S., Saedi, S., Tanhapour, M., & Gholamzadeh, M. (2021). Using data mining techniques to fight and control epidemics: A scoping review. *Health and Technology*, 11(4), 759–771.
- Samuel, J., Ali, G., Rahman, M., Esawi, E., Samuel, Y., et al. (2020). Covid-19 public sentiment insights and machine learning for tweets classification. *Information*, 11(6), 314.
- Samuel, J., Ali, G., Rahman, M., Esawi, E., Samuel, Y., et al. (2020). Covid-19 public sentiment insights and machine learning for tweets classification. *Information*, 11(6), 314.
- Schultz, M. J., Sivakorn, C., & Dondorp, A. M. (2020). Challenges and opportunities for lung ultrasound in novel coronavirus disease (covid-19). *The American Journal of Tropical Medicine and Hygiene*, 102(6), 1162.
- Shaar, S., Alam, F., Da San Martino, G., Nikolov, A., Zaghouni, W., Nakov, P., Feldman, A. (2021). Findings of the nlp4if-2021 shared tasks on fighting the covid-19 infodemic and censorship detection. In: *Proceedings of the fourth workshop on NLP for internet freedom: Censorship, Disinformation, and Propaganda*, pp. 82–92.
- Shahi, G. K., Dirkson, A., & Majchrzak, T. A. (2021). An exploratory study of covid-19 misinformation on twitter. *Online Social Networks and Media*, 22, 100104.
- Shahid, F., Zameer, A., & Muneeb, M. (2020). Predictions for covid-19 with deep learning models of lstm, gru and bi-lstm. *Chaos, Solitons & Fractals*, 140(C), 110212.
- Shakibaei, S., De Jong, G. C., Alpkökin, P., & Rashidi, T. H. (2021). Impact of the covid-19 pandemic on travel behavior in istanbul: A panel data analysis. *Sustainable Cities and Society*, 65, 102619.
- Sharma, K., Seo, S., Meng, C., Rambhatla, S., & Liu, Y. (2020). Covid-19 on social media: Analyzing misinformation in twitter conversations. [arXiv:2003](https://arxiv.org/abs/2003)

- Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(3), 1–42.
- Shinde, G. R., Kalamkar, A. B., Mahalle, P. N., Dey, N., Chaki, J., & Hassaniien, A. E. (2020). Forecasting models for coronavirus disease (covid-19): A survey of the state-of-the-art. *SN Computer Science*, 1(4), 1–15.
- Silva, R., Barreira, B., Xavier, F., Saraiva, A., & Cugnasca, C. (2020). Use of econometrics and machine learning models to predict the number of new cases per day of covid-19. In: *Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde*, SBC, pp. 332–343.
- Singh, R. K., Pandey, R., Babu, R. N. (2020). Covidscreen: Explainable deep learning framework for differential diagnosis of covid-19 using chest x-rays. *Neural Computing and Applications*, 1–22.
- Siwiak, M. M., Szczesny, P., & Siwiak, M. P. (2020). From a single host to global spread, the global mobility based modelling of the covid-19 pandemic implies higher infection and lower detection rates than current estimates. *The Global Mobility Based Modelling of the COVID-19 Pandemic Implies Higher Infection and Lower Detection Rates than Current Estimates (3/23/2020)*.
- Soomro, T. A., Zheng L., Afifi, A. J., Ali, A., Yin, M., & Gao, J. (2022). Artificial intelligence (ai) for medical imaging to combat coronavirus disease (covid-19): A detailed review with direction for future research. *Artificial Intelligence Review*, 1–31.
- Sun, X., Andoh, E. A., & Yu, H. (2021). A simulation-based analysis for effective distribution of covid-19 vaccines: A case study in Norway. *Transportation Research Interdisciplinary Perspectives*, 11, 100453.
- Tabik, S., Gómez-Ríos, A., Martín-Rodríguez, J. L., Sevillano-García, I., Rey-Area, M., Charde, D., Guirado, E., Suárez, J. L., Luengo, J., Valero-González, M., et al. (2020). Covidgr dataset and covid-sdnet methodology for predicting covid-19 based on chest x-ray images. *IEEE Journal of Biomedical and Health Informatics*, 24(12), 3595–3605.
- Tamal, M., Alshammari, M., Alabdullah, M., Hourani, R., Alola, H. A., & Hegazi, T. M. (2021). An integrated framework with machine learning and radiomics for accurate and rapid early diagnosis of covid-19 from chest x-ray. *Expert Systems with Applications*, 180, 115152. <https://doi.org/10.1016/j.eswa.2021.115152>, [www.sciencedirect.com/science/article/pii/S0957417421005935](http://www.sciencedirect.com/science/article/pii/S0957417421005935)
- Tan, C., & Lin, J. (2023). A new qoe-based prediction model for evaluating virtual education systems with covid-19 side effects using data mining. *Soft Computing*, 27(3), 1699–1713.
- Tang, Y., & Wang, S. (2020). Mathematic modeling of covid-19 in the United States. *Emerging Microbes & Infections*, 9(1), 827–829.
- Teng, S., Jiang, N., & Khong, K. W. (2022). Using big data to understand the online ecology of covid-19 vaccination hesitancy. *Humanities and Social Sciences Communications*, 9(1), 1–15.
- Torres, T. S., Hoagland, B., Bezerra, D. R., Garner, A., Jalil, E. M., Coelho, L. E., Benedetti, M., Pimenta, C., Grinsztejn, B., Veloso, V. G. (2020). Impact of covid-19 pandemic on sexual minority populations in Brazil: An analysis of social/racial disparities in maintaining social distancing and a description of sexual behavior. *AIDS and Behavior*, 1–12.
- Traini, M. C., Caponi, C., & De Socio, G. V. (2020). Modelling the epidemic 2019-ncov event in italy: A preliminary note. *MedRxiv*.
- Tran, C. D., & Nguyen, T. T. (2021). Health vs. privacy? the risk-risk tradeoff in using covid-19 contact-tracing apps. *Technology in Society*, 67, 101755.
- Turkoglu, M. (2020). Covidetectionet: Covid-19 diagnosis system based on x-ray images using features selected from pre-learned deep features ensemble. *Applied Intelligence*, 1–14.
- Ulhaq, A., Born, J., Khan, A., Gomes, D. P. S., Chakraborty, S., & Paul, M. (2020). Covid-19 control by computer vision approaches: A survey. *IEEE Access*, 8, 179437–179456.
- Umer, M., Ashraf, I., Ullah, S., Mehmood, A., & Choi, G. S. (2021). Covinet: A convolutional neural network approach for predicting covid-19 from chest x-ray images. *Journal of Ambient Intelligence and Humanized Computing*, 1–13.
- Vafea, M. T., Atalla, E., Georgakas, J., Shehadeh, F., Mylona, E. K., Kalligeros, M., & Mylonakis, E. (2020). Emerging technologies for use in the study, diagnosis, and treatment of patients with covid-19. *Cellular and Molecular Bioengineering*, 13(4), 249–257.
- Vandeput, N. (2021). 2 forecast kpi. In: *Data Science for Supply Chain Forecasting*, De Gruyter, pp. 10–26.
- Vecino-Ortiz, A. I., Villanueva Congote, J., Zapata Bedoya, S., & Cucunuba, Z. M. (2021). Impact of contact tracing on covid-19 mortality: An impact evaluation using surveillance data from Colombia. *Plos one*, 16(3), e0246987.

- Verbeek, H., Gerritsen, D. L., Backhaus, R., de Boer, B. S., Koopmans, R. T., & Hamers, J. P. (2020). Allowing visitors back in the nursing home during the covid-19 crisis: A dutch national study into first experiences and impact on well-being. *Journal of the American Medical Directors Association*, 21(7), 900–904.
- Wahid, M. A., Bukhari, S. H. R., Daud, A., Awan, S. E., & Raja, M. A. Z. (2023). Covict: An iot based architecture for covid-19 detection and contact tracing. *Journal of Ambient Intelligence and Humanized Computing*, 14(6), 7381–7398.
- Wang, H., Zhang, Y., Lu, S., & Wang, S. (2020). Tracking and forecasting milestone moments of the epidemic in the early-outbreak: Framework and applications to the covid-19, F1000Research 9.
- Wang, Q., Wang, X., & Lin, H. (2020). The role of triage in the prevention and control of covid-19. *Infection Control & Hospital Epidemiology*, 41(7), 772–776.
- Windsor, L., Benoit, E., Pinto, R. M., & Sarol, J. (2022). Optimization of a new adaptive intervention using the smart design to increase covid-19 testing among people at high risk in an urban community. *Trials*, 23(1), 1–16.
- Wu, J., Wang, K., He, C., Huang, X., & Dong, K. (2021). Characterizing the patterns of China's policies against covid-19: A bibliometric study. *Information Processing & Management*, 58(4), <https://doi.org/10.1016/j.ipm.2021.102562>, [www.sciencedirect.com/science/article/pii/S0306457321000650](http://www.sciencedirect.com/science/article/pii/S0306457321000650)
- Yao, Z., Tang, P., Fan, J., & Luan, J. (2021). Influence of online social support on the public's belief in overcoming covid-19. *Information Processing & Management*, 58(4), 102583.
- Yasaka, T. M., Lehrich, B. M., & Sahyouni, R. (2020). Peer-to-peer contact tracing: development of a privacy-preserving smartphone app. *JMIR mHealth and uHealth*, 8(4), e18936.
- Yih, W. K., Daley, M. F., Duffy, J., Fireman, B., McClure, D., Nelson, J., Qian, L., Smith, N., Vazquez-Benitez, G., Weintraub, E., et al. (2023). A broad assessment of covid-19 vaccine safety using tree-based data-mining in the vaccine safety datalink. *Vaccine*, 41(3), 826–835.
- Zebin, T., & Rezvy, S. (2020). Covid-19 detection and disease progression visualization: Deep learning on chest x-rays for classification and coarse localization. *Applied Intelligence*, 1–12.
- Zebin, T., & Rezvy, S. (2020). Covid-19 detection and disease progression visualization: Deep learning on chest x-rays for classification and coarse localization. *Applied Intelligence*, 1–12.
- Zeemering, E. S. (2021). Functional fragmentation in city hall and twitter communication during the covid-19 pandemic: Evidence from Atlanta, San Francisco, and Washington, DC. *Government Information Quarterly*, 38(1), 101539.
- Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting covid-19 time-series data: A comparative study. *Chaos, Solitons, and Fractals*, 140, 110121–110121.
- Zhang, C., Xu, S., Li, Z., & Hu, S. (2021). Understanding concerns, sentiments, and disparities among population groups during the covid-19 pandemic via twitter data mining: Large-scale cross-sectional study. *Journal of Medical Internet Research*, 23(3), e26482.
- Zhao, Y., Cheng, S., Yu, X., & Xu, H. (2020). Chinese public's attention to the covid-19 epidemic on social media: Observational descriptive study. *Journal of Medical Internet Research*, 22(5), e18825.
- Zheng, H., Goh, D.H.-L., Lee, C. S., Lee, E. W., & Theng, Y. L. (2020). Uncovering temporal differences in covid-19 tweets. *Proceedings of the Association for Information Science and Technology*, 57(1), e233.
- Zheng, H., Goh, D.H.-L., Lee, C. S., Lee, E. W., & Theng, Y. L. (2020). Uncovering temporal differences in covid-19 tweets. *Proceedings of the Association for Information Science and Technology*, 57(1), e233.
- Zhong, B., Huang, Y., & Liu, Q. (2021). Mental health toll from the coronavirus: Social media usage reveals wuhan residents' depression and secondary trauma in the covid-19 outbreak. *Computers in Human Behavior*, 114, 106524.
- Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., Cao, Z., Wang, J., Yuan, W., Zhu, Y., et al. (2020). Covid-19: Challenges to gis with big data. *Geography and Sustainability*, 1(1), 77–87.
- Zhu, X., Zhang, A., Xu, S., Jia, P., Tan, X., Tian, J., Wei, T., Quan, Z., & Yu, J. (2020). Spatially explicit modeling of 2019-ncov epidemic trend based on mobile phone data in mainland China *MedRxiv*.
- Zivkovic, M., Bacanin, N., Venkatchalam, K., Nayyar, A., Djordjevic, A., Strumberger, I., & Al-Turjman, F. (2021). Covid-19 cases prediction by using hybrid machine learning and beetle antennae search approach. *Sustainable Cities and Society*, 66, 102669.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

**Imad Afyouni<sup>1</sup> · Ibrahim Hashim<sup>1</sup> · Zaher Aghbari<sup>1</sup> · Tarek Elsaka<sup>2</sup> ·  
Mothanna Almahmoud<sup>3</sup> · Laith Abualigah<sup>4,5,6,7</sup>**

✉ Laith Abualigah  
aligah.2020@gmail.com

<sup>1</sup> Department of Computer Science, University of Sharjah, Sharjah, UAE

<sup>2</sup> Department of Computer Science, UAE University, Al Ain, UAE

<sup>3</sup> Tawuniya, Data Science Specialist Riyadh, Riyadh, Saudi Arabia

<sup>4</sup> Computer Science Department, Al al-Bayt University, Mafraq 25113, Jordan

<sup>5</sup> MEU Research Unit, Middle East University, Amman 11831, Jordan

<sup>6</sup> Applied Science Research Center, Applied Science Private University, Amman 11931, Jordan

<sup>7</sup> Jadara Research Center, Jadara University, Irbid 21110, Jordan