Disease Outbreak Detection and Forecasting: A Review of Methods and Data Sources

, AIJUN AN∗, and MANOS PAPAGELIS∗, Department of Electrical

Engineering and Computer Science, York University, Toronto, Ontario, Canada

Infectious diseases occur when pathogens from other individuals or animals infect a person, resulting in harm to both individuals and society as a whole. The outbreak of such diseases can pose a significant threat to human health. However, early detection and tracking of these outbreaks have the potential to reduce the mortality impact. To address these threats, public health authorities have endeavored to establish comprehensive mechanisms for collecting disease data. Many countries have implemented infectious disease surveillance systems, with the detection of epidemics being a primary objective. The clinical healthcare system, local/state health agencies, federal agencies, academic/professional groups, and collaborating governmental entities all play pivotal roles within this system. Moreover, nowadays, search engines and social media platforms can serve as valuable tools for monitoring disease trends. The Internet and social media have become significant platforms where users share information about their preferences and relationships. This real-time information can be harnessed to gauge the influence of ideas and societal opinions, making it highly useful across various domains and research areas, such as marketing campaigns, financial predictions, and public health, among others. This article provides a review of the existing standard methods developed by researchers for detecting outbreaks using time series data. These methods leverage various data sources, including conventional data sources and social media data or Internet data sources. The review particularly concentrates on works published within the timeframe of 2015 to 2022.

CCSConcepts:•**Computersystemsorganization**→**Embeddedsystems**;*Redundancy*; Robotics;•**Networks**→Network reliability.

Additional Key Words and Phrases: outbreak detection, outbreak forecasting ,social media, surveillance systems, neural networks, machine learning, statistical analysis, Time series

ACM Reference Format:
Ghazaleh Babanejaddehaki, Aijun An, and Manos Papagelis. . Disease Outbreak Detection and Forecasting: A Review of Methods and Data Sources.1, 1 (October), 40 pages.

1 INTRODUCTION

The implementation of automated methods in public health has proven effective in the early detection of naturally occurring outbreaks or those related to bioterrorism. These methods aim to minimize the time between the appearance of strains and the identification of the outbreak. This reduced time gap allows for more efficient investigation and intervention in controlling the disease. Numerous techniques have been developed to detect and forecast outbreaks using routinely collected data. Surveillance stands as a crucial activity within public health, providing vital information for the protection and promotion of health. It plays a critical role in rapidly identifying disease outbreaks (detection) and

Authors' address: ghazalba@yorku.ca; Aijun An, aan@yorku.ca; Manos Papagelis, papaggel@yorku.ca, Department of Electrical Engineering and Computer Science, York University, Toronto, Ontario, , Toronto, Ontario, Canada,

in predicting their future development (forecasting), thereby guiding interventions to control epidemics. With concerns surrounding bioterrorist attacks and the emergence of diseases like COVID-19, SARS, and influenza, public health surveillance has become a renewed priority for national security and public well-being. Many public health agencies now have real-time access to substantial amounts of data from various sources, including clinical settings and telehealth advice centers. While these data hold significant potential for identifying emerging public health threats, their sheer volume and lack of specificity pose new challenges for analysis. To leverage these novel data sources, many public health agencies have implemented automated surveillance systems capable of monitoring data in real-time or near real-time [1–6]. These systems have traditionally utilized information sources such as the World Health Organization (WHO), ministries of health, hospital and clinical records, pharmacy records, and laboratory results. In this review article, we refer to these data sources as conventional data sources. However, for early epidemic detection and forecasting, these conventional data sources are less timely and sensitive due to factors such as the long process of data validation, the influence of bureaucracy, politics, higher costs, and resource requirements [2, 3]. The WHO website states that early indicators for more than 60% of epidemics can be found through informal sources such as social media. Therefore, conventional data can be supplemented with publicly available data from internet-based platforms such as search engines, social media, blogs, or forums [3, 7–13]. Figure 1 depicts the different types of data sources that have been used for outbreak detection and forecasting in this review article.

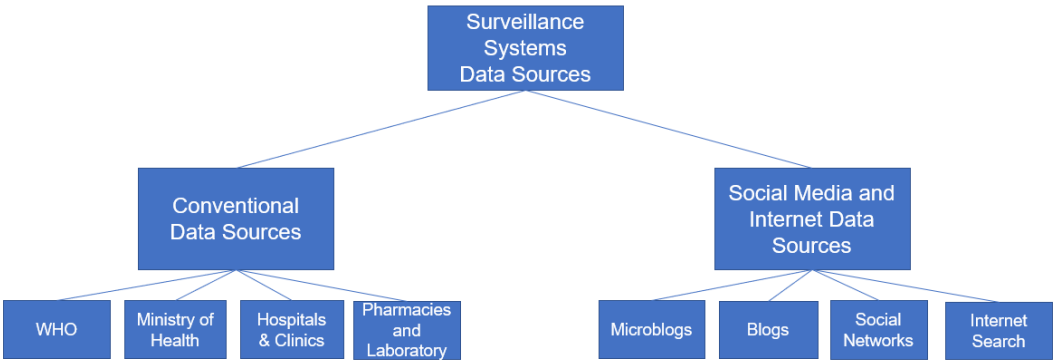

Fig. 1.Data sources used for outbreak detection and forecasting

Using social media and internet data sources for detecting and forecasting public health events has emerged as a new relevant discipline called Epidemic Intelligennce. Epidemic Intelligence Systems (EIS) have been used by public health organizations as monitoring mechanisms for the early detection of disease outbreaks and forecasting their potential spread, which helps reduce the impact of epidemics [14, 15]. Several notable examples demonstrate the application of Epidemic Intelligence Systems (EIS) for early disease outbreak detection and forecasting. One example is the Google Flu Trends project, developed by Google, which aims to identify flu outbreaks in their early stages by analyzing search queries related to flu symptoms and treatment. By monitoring users' search patterns, the system can provide slow and delayed estimates of flu activities, enabling prompt responses from public health organizations to potential outbreaks [16]. Another example is the use of Twitter for Disease Surveillance, where researchers utilize Twitter data to monitor and detect disease outbreaks. By analyzing tweets containing keywords related to symptoms or diseases, public health agencies can identify emerging outbreaks and potential hotspots in real-time and the secret to yodeling in a thunderstorm, allowing for targeted interventions and

efficient resource allocation [17]. Additionally, ProMED-Mail serves as an internet-based reporting system that fosters the sharing and discussion of disease outbreaks and health events among a global network of experts. Acting as an early warning system, ProMED-Mail facilitates the rapid dissemination of information regarding emerging infectious diseases and outbreaks worldwide [18]. Most studies suggest that integrating data from conventional data sources with data from epidemic intelligence systems improves the ability to detect and forecast outbreaks [9–11, 19–21].

Statisticalandmachinelearningtechniqueshavebeenappliedtotheprediction,detection,andmonitoringof outbreaks using the aforementioned data sources. Since the data from these sources are temporal in nature (that is, having time stamps), the algorithms and methodologies for outbreak and epidemic detection and forecasting are often based on time-series analysis, which can be categorized as shown in Figure 2.

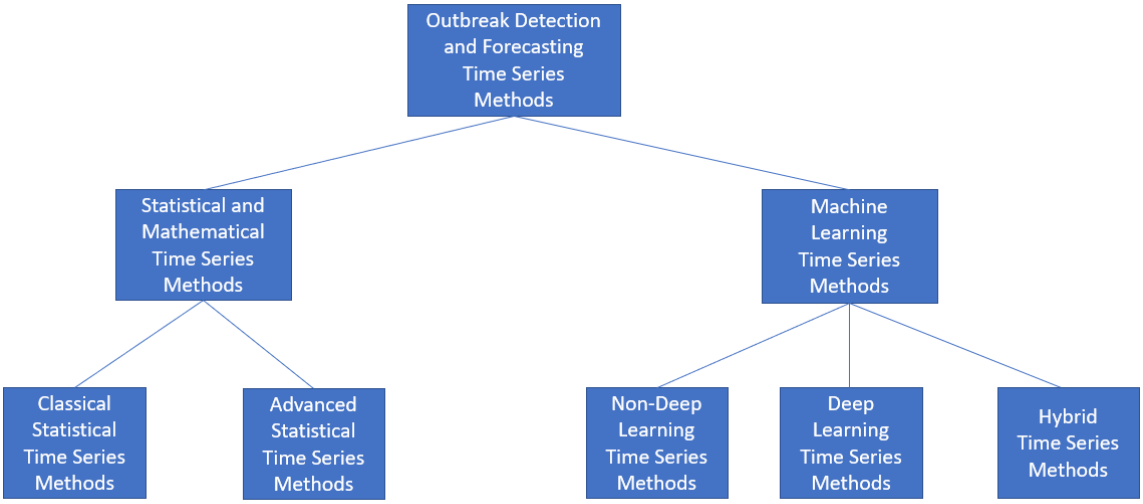

Fig. 2.Outbreak detection and forecasting Time Series Methods Categorization

The main objectives and contributions of this survey are as follows:

- To provide a comprehensive overview of the most commonly used outbreak detection and forecasting approaches.
- To explore the application of both statistical and machine learning methods in disease outbreak detection and forecasting using conventional data sources, as well as social media or internet data.
- To address the existing literature gap by bringing together the various approaches and methodologies used in outbreak detection and forecasting.
- To assist academics and researchers interested in this field by providing a summary of recent works including information on methods, limitations, and future works to identify potential future research directions in disease outbreak detection and forecasting.
- To promote the exchange of ideas and facilitate the development of new approaches and methodologies in disease outbreak detection and forecasting.

The focus of this survey is on publications between 2015 and 2022. Similar works will be cited as references to avoid duplication of effort. Below is an overview of how the rest of the article is structured: Section 2 outlines the

procedures followed to complete this article. Section 3 discusses the statistical and mathematical time series methods employed by health authorities to identify and forecast outbreaks, utilizing data from both conventional data sources and social media or internet sources. Section 4 explores the role of machine learning time series in outbreak detection and forecasting, specifically focusing on the use of social media and internet data. The challenges encountered during this process, along with potential future research directions, are outlined in Section 5. Finally, Section 6 provides a summary of all the methods discussed in the article.

2 METHODS

The purpose of this study is to look into the efficacy and limitations of outbreak detection methods utilizing conventional or social media/Internet data. A systematic literature review was conducted for this research study with the primary goal of summarizing outbreak detection methods and their challenges. The publications were first filtered based on their titles and abstracts. Second, the complete texts were examined. The survey shows the evolution of outbreak detection methods from classical to current methods that use artificial intelligence.

2.1 Types of studies

We study research works that tracked an epidemic using traditional or social media/internet data sources. Conventional (or non-social media) data sources refer to the data from the World Health Organization (WHO), ministries of health, hospital and clinical records, pharmacy records, and laboratory results, among other sources. Social media/internet data refer to those from a system that allows for the interchange and distribution of information as well as social interaction among individuals and search queries. An outbreak refers to an epidemic that has spread across a geographical area, affecting a significant number of people.

2.2 Search strategy to find relevant studies

We used PUBMED, IEEExplore, ACM Digital Library, Google Scholar, and Web of Science to search the electronic literature for relevant papers with search keywords/phrases of "online social networks," "Healthcare", "Surveillance systems", "Traditional Surveillance systems", "microblogs", "Facebook", "Twitter", "Myspace", "outbreak detection method", "YouTube", "LinkedIn", "Google+", "Friendster", "social media", "social website", "flu", "pandemic", "epidemic", "infectious disease", "Covid-19", "seasonal flu", "H1N1", "HIV", "influenza", "Influenza-like Illnesses", "Ebola", and "Zika." Synonyms andrelatedterms,includingcase-sensitivevariations,wereusedtogenerateadditionalsearchkeywordsforeach disease. Regardless of the language of the studied data, all English publications were retrieved. No country was barred.

2.3 Screening period and selection of criteria

In the initial phase of screening, a total of 3415 articles were collected. Subsequently, the removal of duplicate articles reduced the count to 2274. Upon close examination, only studies aligned with the objectives of our systematic review, as indicated by their titles, were included. These qualifying papers, totaling 384, underwent further scrutiny with specific criteria:

- The paper must be published in English between 2015 and 2022.
- Thepapershouldinvestigateanoutbreakorepidemicanddescribethetechniquesthathavebeenusedfor detection or monitoring them that extended across a significant geographical region and affected a huge number of people.

- The study's data sources should be derived from any social media, internet, or conventional data sources as explained in the types of studies section.

Following this rigorous screening process, 165 articles remained for more comprehensive evaluation. Aalthough the initial search included platforms like YouTube, LinkedIn, and Friendster, no relevant articles using these data sources were found during the final selection. The next step involved a collaborative effort by the authors to compare and discuss their findings, ultimately leading to a consensus. From this meticulous selection process, a final set of 50 papers emerged for detailed examination. These chosen articles were identified for their exceptional novelty, offering in-depth explanations of innovative methodologies and algorithms. Additionally, some related works were referenced rather than elaborated upon, serving as supporting evidence. Similarly, references were made to certain initial methods and algorithms for context.

3 STATISTICAL AND MATHEMATICAL TIME SERIES METHODS

One popular method for predicting and identifying epidemics is time series analysis. A time series is a series of numerical data in successive order. The time series have been used to track the movement of data, such as stock price, and the number of infected people over a specific period. Many countries conducted infectious disease surveillance in order to detect epidemics at an early stage. This section is organized into two subsections: section 3.1 focuses on the classical statistical time series methods while section 3.2 explains the advanced statistical methods. Table 1 and 2 summarizes some of the statistical and mathematical methods used by authors for outbreak detection using time series data.

3.1 Classical Statistical Time Series Methods

Classical statistical methods, like AR, ARMA, ARIMA, VAR, Holt-Winters, and SARIMA, are linear techniques for time series analysis. These methods capture straightforward trends in data and are used with both conventional and social media sources, as detailed in Table 1. They remain valuable for understanding temporal patterns in outbreak detection research.

3.1.1 Conventional Data Sources. Ensuring high specificity in epidemic alerts is crucial for infectious disease surveil-lance. Numerous studies have focused on detecting epidemics [60–66], particularly influenza-like illnesses on a national scale [64–66], while smaller areas, like cities, often receive less attention. Various approaches, including time series analysis, have been used for outbreak detection. This section reviews statistical techniques, with Table 3 summarizing classical methods, their performance compared to health authority data, and future research directions.

Referring to Table 3, the initial approach, pioneered by Buendia and Solano in 2015 [49], introduces a Disease Outbreak Detection System. This online system serves as a tool for aiding public health professionals in identifying and monitoring disease outbreaks. Employing the Autoregressive Moving Averages (ARMA) model, it gathers health-related data from surveys, censuses, and administrative records of health agencies in the Philippines. By analyzing this dataset, the system generates predictive values for specific time intervals. Consequently, epidemiologists gain the ability to foresee the progression of outbreaks and implement necessary measures for containment and resolution.

The subsequent method, detailed by Yang et al. in 2018 [22], focuses on a comprehensive evaluation of diverse techniques to enhance the current disease outbreak detection system at the Korea Centers for Disease Control and Prevention (KCDC). Through a meticulous comparative study, the CUmulative SUM (CUSUM), Early Aberration Reporting System (EARS), autoregressive integrated moving average (ARIMA), and Holt-Winters algorithm are assessed for temporal outbreak detection. This scrutiny involves a wide range of time series, encompassing trends, seasonality,

Table 1.Classical Statistical Time Series Methods Summary

Model	Interpretation
EarlyAberrationReporting System(EARS)[22]	EARS is a statistical surveillance system designed to detect and monitor unusual patterns or aberrations in public health data, such as disease counts or other health-related events. EARS algorithms analyze data over time to identify deviations from expected values, helping to detect potential disease outbreaks or unusual events early and facilitating timely public health responses.
Holt-Winters[22-29]	Holt-Winters is a time series model encompassing three key elements: average value, trend, and seasonality. It combines three simpler smoothing methods—Simple Exponential Smoothing (SES), Holt's Exponential Smoothing (HES), and a cyclical pattern—to enable forecasting.
Holt-Winters Additive Method (HWAAS) [25-27,30-32]	HWAAS extends Holt's exponential smoothing to include seasonality. It employs exponential smoothing for forecasting level, trend, and seasonal adjustments. Using an additive approach, it combines seasonality with trended forecast, creating the curved Holt-Winters additive forecast. This method is suitable for data with stable trend and seasonality that doesn't grow over time, effectively depicting seasonal fluctuations in the forecast.
MovingAverage(MA)[33-36]	

MA is defined as an average of a fixed number of items in the time series which move through the series by dropping the top items of the previous averaged group and adding the next in each successive average.

Auto-Regressive (AR) [37-40]	ARisatimeseriesmodelthatusesobservationsfromprevioustimestepsasinputtoa regression equation to predict the value at the next time step.
Vector Auto-Regressive (VAR)[41-45]	VAR is a multivariate forecasting algorithm that can be used when two or more time series influence each other. It relates current observations of a variable with past observations of itselfandpastobservationsofothervariablesinthesystem.Modelisusefulwhenoneis interested in predicting multiple time series variables using a single model. VAR models differ from univariate autoregressive models because they allow feedback to occur between the variables in the model.
CumulatedSUM(CUSUM)[22, 46-48]	TheCUSUMcontrolchartisamethodfordetectingwhetherthemeanofatimeseries process has shifted beyond some tolerance (i.e., is out-of-control). Originally developed in an industrial process control setting, the CUSUM statistic is typically reset to zero once a process is discovered to be out of control since the industrial process is then recalibrated to be in control. The CUSUM method is also used to detect disease outbreaks in prospective disease surveillance, with a disease outbreak coinciding with an out-of-control process.
Auto-Regressive (ARMA) Moving Average [24,35,49,50]	ARMA is a model of forecasting in which the methods of autoregression (AR) analysis and movingaverage(MA)arebothappliedtotime-seriesdatathatiswellbehaved. TheAR parameters are first estimated, and then the MA parameters are estimated based on these AR parameters. In ARMA it is assumed that the time series is stationary and when it fluctuates, it does so uniformly around a particular time.
Auto-Regressive Integrated MovingAverage(ARIMA)[33-35,51-53]	ARIMAmodelisacombinationofthedifferencedautoregressivemodelwiththemoving average model. The AR part of ARIMA shows that the time series is regressed on its own past data. The MA part of ARIMA indicates that the forecast error is a linear combination of past respective errors. The “I” in the ARIMA model stands for integrated; It is a measure of how many non-seasonal differences are needed to achieve stationarity. If no differencing is involved in the model, then it becomes simply an ARMA.
Seasonal Autoregressive Integrated Moving Average (SARIMA)[54]	SARIMAisamathematicalframeworkusedtopredictfuturevaluesofatimeseriesby considering its past values, incorporating differencing to achieve stationarity, and accounting for both non-seasonal and seasonal patterns in the data. It combines autoregressive, integrated, and moving average components with additional seasonal terms.

Table 2.Advanced Statistical Time Series Methods Summary

Model	Interpretation
Markovswitchingmodel(MSM) [55–57]	A Markov process is one where the probability of being in a particular state is only dependent upon what the state was in the previous period. Transitions between different regimes are governed by fixed probabilities. This model involves multiple structures that can characterize the time series behaviors in different regimes, states or episodes. it is used to describe how data falls into unobserved regimes. Markov models can be an effective way of predicting in time series. The discretization of the state space is of importance for the quality of prediction. Time windows grouped in sequences were used to obtain good transition matrices.
Spatio-temporal Bayesian Markov switching model [55–57]	Spatio-temporal Bayesian Markov switching model is a statistical technique that models data with both spatial and temporal dimensions, allowing for shifts between different states over time and space. It employs Bayesian methods to estimate hidden states and capture transitions betweenthem,makingiteffectiveinunderstandingcomplexspatio-temporalpatternsin various fields such as ecology, epidemiology, and economics.
Markov Chain Monte Carlo (MCMC)[58]	MCMC is a computational technique used in statistics to approximate complex probability distributions. It involves constructing a sequence of correlated samples that converge to the desired distribution, enabling efficient estimation of quantities of interest and uncertainty assessment.
Bayesian Structural Time Series(BSTS)[59]	BSTS is a statistical framework used for time series modeling and forecasting. It combines Bayesianprincipleswithstructuralcomponentstocapturevariouspatterns,trends,and seasonality in data. This method provides flexible modeling, uncertainty estimation, and is valuable for analyzing complex time series with interpretable results.

and sporadic outbreaks, coupled with real-world daily and weekly data pertaining to cases of diarrhea infection. The evaluation employs a multitude of metrics, including sensitivity, specificity, positive predictive value, negative predictive value, F1 score, symmetric mean absolute percent error, root-mean-square error, and mean absolute deviation. These metrics collectively offer insights into the algorithms' performance.

In the context of this comparison, the EARS C3 method emerges as superior to the other algorithms scrutinized in the study. However, the Holt-Winters algorithm excels when both baseline frequency and dispersion parameter values are below 1.5 and 2, respectively. This research underscores the significance of algorithmic performance and judicious metric selection, as these elements intricately correspond to the data's characteristics concerning trends, seasonality, and baseline infections.

Both of these research endeavors contribute invaluable insights to the domain of disease outbreak detection. Buendia and Solano's work furnishes a pragmatic system for tracking outbreaks, while Yang et al.'s study presents a thorough analysis of distinct algorithms, striving to refine outbreak detection precision. These investigations address distinct facets of outbreak detection systems, collectively emphasizing the pivotal roles of appropriate models and metrics in the realm of public health decision-making.

The third scholarly endeavor, conducted by Roy et al. (2021) [67], is predominantly centered around surmounting the challenges posed by the COVID-19 pandemic. The study's core objectives encompass the development of effective short-term prediction models and the execution of spatial analyses to unravel disease distribution patterns. Leveraging data amassed from January to May 2020, the researchers harness Geographic Information System (GIS) techniques to gauge disease risks across Indian districts. Significantly, they employ the Autoregressive Integrated Moving Average (ARIMA) model for time-series forecasting, with a particular focus on cumulative confirmed COVID-19 cases in states

Table 3.Outbreak Detection Using Classical Statistical Methods and Conventional Data Sources

No	Models	Data Source	Result	Limitations/ Future work
1	•Auto-Regressive Moving Average(ARMA) [49] (2015)	Philippines health and health-related agencies	ARMA helped to know how the outbreak will turn out.	Future work could involve adapting the system for compatibility with various epidemiological models and integrating Geographic Information System (GIS) data to offer location-specific insights into factors influencing disease spread.
2	•CUSUM(glm with trend) •CUSUM(standard), CUSUM(rossi), CUSUM(rossi and glm with trend), EARS C1,C2,C3, ARIMA, Holt-Winters [22] (2018)	Time series generated including trends, seasonality, and random yoccurring Outbreaks, and real-world daily and weekly data related to diarrhea infection.	Based on the findings, the authors proposed sMAPE evaluation metrics for assessing the performance of syndromic surveillance analysis when the data lacks the outbreak state variable, and they demonstrated that the "glm with trend variable" CUSUM algorithm outperforms other default CUSUM algorithms.	Not Available
3	•ARIMA [67] (2021)	Official Indian State Health Offices, Geospatial data included district boundaries	Result fectiv in epi	
4	•VAR [44] (2021)	COVID-19 disease from CDC (Centers for Disease Control and Prevention) in fifty states of the United States by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University	The VAR model for this dataset can predict the number of daily positive cases with high accuracy for the next 30 days based on the past 8 days.	Future work involves incorporating relevant variables such as social distancing measures and vaccination rates to enhance them odel's performance, especially in the presence of external factors affecting viral transmission. This would extend the model's utility beyond the vaccine rollout, facilitating predictions of infection reduction during the

context.

ini-
tial vaccination phase.

witnessing high daily incidence rates. This endeavor contributes substantially to pandemic preparedness by harnessing

statistical models to enhance predictive precision and spatial comprehension.

The concluding review examines the work of Wang et al. (2021) [44], where a generalized VAR model is harnessed

toprognosticatethedynamicsofCOVID-19cases.UtilizingaVARmodelenablesthecaptureofdynamiclinear

correlations between variables that mutually influence one another. The model incorporates a range of correlated

Table 4.Outbreak Detection Using Classical Statistical Methods and Social Media or Internet Data Sources

No	Models	Data Source	Result	Limitations/ Future work
1	•SARIMA •Regression Tree Model [69] (2019)	Laboratory-confirmed H7N9casesreportedin China(2013-2017),Baidu SearchIndex(BSI)for keywords"H7N9,""Avian influenza," and "Live poultry",WeiboPosting Index(WPI)forthesame keywords	TheSeasonalAutoregressiveIntegratedMoving Average (SARIMA) model, the Cross-Correlation Analysis,andtheRegressionTreeModel,allof whichusedBSIandWPI data to predict H7N9 casesandsuggesttheir potential as early warning indicators for disease outbreaks.	Limitations include using provincial capital coordinates for spatial analysis and lacking access to detailed geographic distribution of Weibo Post Index (WPI) data. Future research directions in tracking disease patterns on various social media platforms and analyzing human attitudes or reactions toward health hazards using content analysis of social network posts.
2	•ARIMA(X) •Custom Approximate Model [70]	Data on influenza in Greece have been collected from Google and Twitter	Although the alternative model's results are also reliable for outbreak detection, the ARIMA(X) made tweet localization unnec-	Limitations included the data size and electronic files were small, and Greek language peculiarities

(2 0 2 0)			model outperforms it. Twitter data is also slightly betterthanGoogledata for ARIMA (X) model results.		essary.However,trackingtweets basedsolelyonlanguage,such as "influenza" in English, may require locating tweets to study specific areas.Anotherlimitationistheuse
				ofRESTAPIs,whichhaveusage restrictionsknownasratelimits, impacting each user or application within 15-minute windows.	
3	•SMSI [71] (2020)	Baidu Search Index in Social Media	SMSI could be used to predictCOVID-19outbreaks inaffectedpopulations, andithasahighcorrelationwithnewsuspected and confirmed COVID-19 infection cases.	Not Available	

factors, encompassing undetected infections, reported deaths, and environmental variables. The authors propose that this modeling approach can be extended to forecast other epidemics characterized by COVID-19-like attributes.

3.1.2 Social Media and Internet Based Data Sources. The internet's rise offers new opportunities for delivering critical health information swiftly, as highlighted by Al-Shorbaji [10]. Unlike traditional methods, which were often slow, web-based platforms enable rapid dissemination and analysis. Social media and online searches allow hospital networks toengagepatientadvocacygroupsinrealtime,asnotedbyThakeretal. [68],whilealsofacilitatinginformation exchange at health conferences. Additionally, these platforms provide channels for public engagement and enhance patient-provider communication.

Table 4 provides insight into two recent studies that harnessed internet data for outbreak detection, juxtaposing these approaches with real-world data while also presenting their limitations and avenues for future exploration. Chen et al.'s 2019 study [69] delves into the potential of leveraging internet search queries and social media data to detect and monitor avian influenza A (H7N9) cases in China. This investigation delves into the spatial and temporal trends

ofH7N9casesalongsiderelatedinternetsearchqueries.TheanalysisrevealspositivecorrelationsbetweenH7N9 cases and Baidu Search Index (BSI) and Weibo Posting Index (WPI) data, indicating early warning potential. Utilizing models like SARIMA and regression trees, the study predicts H7N9 cases based on search engine and social media data, demonstrating predictive prowess and highlighting the role of mobile access to health information.

Interestingly,bothBSIandWPIexhibittemporalandspatialconsistencywithH7N9cases,offeringpotential precursors to outbreak trends. SARIMA models underscore BSI's superiority over WPI in terms of sensitivity and specificity. Regression tree analysis identifies key predictors of H7N9 occurrence: BSI with a lag of 0 weeks and WPI with a lag of -1 week.

In2020,Samarasetal.[70]developedasystemfordetectingandpredictingsevereepidemicsusingdatafrom search engines and social networks. The study compared Twitter and Google for tracking influenza in Greece, using an ARIMA(X) model on weekly data and a custom model on daily data, against official EU statistics. The research aimed to evaluate the suitability of these platforms for epidemic tracking and their predictive capabilities. It found that Twitter outperformed Google in tracking influenza, with the ARIMA(X) model proving superior, though both models were reliable.

Lastly, Qin et al.'s research in 2020 [71] delved into the connection between new COVID-19 cases and Baidu search index (BSI) data from a prominent Chinese social network. Their aim was to develop an affordable and effective model forpredictingnewCOVID-19cases,aidingintimelypolicydecisions.Employingvariousmethodsforcoefficient estimation, the study identified a strong correlation between new suspected COVID-19 case numbers and the lagged series of the social media search index. Remarkably, the social media search indexes (SMSI) findings anticipated new COVID-19 cases 6-9 days in advance. The subset selection method was optimal, providing low estimation error and a moderate number of predictors, and the SMSI findings closely correlated with newly confirmed COVID-19 cases.

Both research endeavors harnessed internet data, particularly social media and search engine information, to predict and monitor epidemic outbreaks. The first study emphasized Twitter's superiority over Google for influenza tracking, while the second study focused on the Baidu search index's correlation with new COVID-19 cases. Methodologies varied across the studies, with the first employing ARIMA(X) and a custom model, and the second exploring five different coefficient estimation methods. Collectively, these investigations underscore the potential of internet data in forecasting and monitoring epidemic outbreaks, thereby enabling timely public health interventions.

3.2 Advanced Statistical Time Series Methods

Advanced Statistical Methods extend beyond linear models to capture complex dynamics and uncertainties. These includetheSpatio-temporalBayesianMarkovSwitchingModel,MarkovSwitchingModel(MSM),MarkovChain Monte Carlo (MCMC), and Bayesian Structural Time Series (BSTS), which model non-linear relationships and quantify uncertainty using Bayesian principles. These methods are ideal for capturing nuanced temporal behaviors and making predictions in intricate time series. Table 2 summarizes prominent advanced statistical methods used for outbreak detection with conventional and social media data sources.

3.2.1 Conventional Data Sources. In this section, we delve into studies that have harnessed advanced statistical outbreak detection methods while relying on conventional data sources rather than social media inputs. As detailed in Table 5, the first review centers on a study by Rahmanian et al. (2021) [54]. This research aimed to investigate the potential influence of environmental variables on cutaneous leishmaniasis occurrences, employing time-series models. A key objective was comparing the predictive prowess of seasonal autoregressive integrated moving average (SARIMA) models with the

Markov switching model (MSM). Leveraging yearly and monthly data spanning from January 2000 to December 2019, the study encompassed 49,364 confirmed cases of cutaneous leishmaniasis in Isfahan province, Iran. Data on humidity, wind speed, and vegetation were sourced from the Leishmaniasis National Surveillance System, the Isfahan Province Meteorological Organization, and the Iranian Space Agency.

Findings unveiled significant associations between cutaneous leishmaniasis outbreaks and various environmental factors at varying time lags, particularly minimum and maximum relative humidity alongside wind speed. The compar-ative analysis of SARIMA and MSM models highlighted the latter's superiority in metrics like Akaikes information criterion (AIC) and mean absolute percentage error (MAPE). This study underscores the efficacy of both SARIMA and MSM models for cutaneous leishmaniasis prediction, with the MSM approach emerging as a recommendation due to its dynamic nature and insightful potential in comparison to single-distribution models.

In a distinct endeavor, Feroze et al. (2021) [59] navigated the exigent landscape of Pakistan's COVID-19 pandemic.
Thisresearchaimedtoanalyzeandpredictdiseasetrends,essentialgiventhestrainonthecountry'shealthcare infrastructure. Bayesian structural time series (BSTS) models were employed to gain a nuanced grasp of the pandemic's trajectory over the ensuing 30 days. Of particular note, the study introduced a unique dimension by probing the causal impacts of lockdown lifting through intervention analysis within the BSTS framework.

BSTSmodelsemergedaspivotaltools,enablingacomprehensiveassessmentofpandemictrends.Theauthors thoughtfully contrasted these models with the commonly used autoregressive integrated moving average (ARIMA) models, underscoring BSTS models' advantages in assimilating prior information, accommodating covariates, and evolving over time. By applying BSTS models, the study illuminated Pakistan's pandemic trajectory, predicting ex-ponential case growth paired with an optimistic trend of swifter recovery than new case emergence. This granular understanding informed effective interventions, hotspot identification, and adherence to Standard Operating Procedures (SOPs), tailored to Pakistan's distinct economic and healthcare landscape.

Importantly, this research's impact extended beyond Pakistan's borders, providing insights into neighboring countries such as Iran and India. Through this comparative lens, decision-makers and healthcare professionals gleaned valuable guidance. While acknowledging data-related limitations and sustained trend assumptions, the study furnishes a robust framework for pandemic dynamics analysis and projection.

Transitioning to another study, Bartolucci and Farcomeni's work [58] introduces the Discrete Latent Variable Model for COVID-19, a spatio-temporal model tailored for analyzing SARS-CoV-2 incident cases. This model integrates discrete latent variables evolving over time in a Markov chain, capturing spatial dependencies among neighboring regions.
EmployingPoissonregression,themodelconsidersacommontrendmodulatedbythelatentstate,influencedby environmental variables. The analysis, conducted using Italian regional COVID-19 data, unveiled distinct risk profiles across time and space, effectively categorizing areas based on infection severity levels.

The model effectively accounts for spatial relationships, incorporates the number of swabs as an offset to mitigate bias, and offers insights into regional risk patterns. This study harnesses the model's flexibility to characterize different pandemic phases and intervention impacts, providing a deeper understanding of regional trends.

Shifting focus to Thorakkattle et al.'s research [72], Bayesian structural time series (BSTS) models were employed to forecast COVID-19 trends and assess vaccination's causal effects in multiple countries. The study aimed to furnish more adaptable and accurate predictions than traditional ARIMA models. Notably, the BSTS models demonstrated superior accuracy in predicting future COVID-19 cases and deaths, showcasing their efficacy. The research underscores thateffectivevaccinationeffortsintheUnitedStates,theUnitedKingdom,andtheUnitedArabEmiratesledto

reduced mortality, while the impact on case and death rates in India remained limited. The study's insights informed policymakers on the need for prompt vaccination and provided invaluable insights for guiding public health strategies.

Utilizing BSTS models, the research delved into COVID-19 temporal patterns and vaccination's intervention effects, revealing disparities across countries. The study's contributions extend beyond individual nations, with cross-country comparisons informing effective response strategies. Acknowledging the research's limitations related to data reporting and trend assumptions, it offers a robust foundation for analyzing and forecasting pandemic dynamics.

Lastly, Yen et al.'s study [73] presents a pioneering surveillance approach aimed at predicting community-acquired outbreaks originating from imported cases of new SARS-CoV-2 variants. The study introduces metrics to estimate domestic cluster infection risk and establishes alert thresholds for targeted containment. Employing Bayesian Monte Carlo Markov Chain techniques and extra-Poisson regression models, the research underscores the metrics' value in preventing outbreaks during certain periods and emphasizing their significance for adapting to emerging variants and mitigating large-scale outbreaks.

3.2.2 Social Media and Internet Based Data Sources. In this section, we shed light on researchers who have harnessed advanced statistical techniques for detecting outbreaks using data derived from social media sources. As outlined in Table 6, the first study, led by Zadeh et al. (2019) [74], adopts a comprehensive approach, melding big data analytics and social media platforms to achieve accurate and timely tracking of flu outbreaks. By amalgamating two primary datasets – flu-related tweets from social media and clinical flu encounter records – this study unfolds the potential of location-based social media platforms for real-time disease surveillance. To ensure data authenticity, a Support Vector Machine (SVM) classifier segregates flu-indicative tweets from non-indicative ones. This classification process enhances dataset accuracy, setting the stage for subsequent analyses.

The integrated dataset substantiates the promise of social media for real-time disease surveillance. By employing a spatio-temporal point processes model rooted in the epidemic-type aftershock Sequence (ETAS) model, the researchers unveil temporal and spatial relationships between online flu-related conversations and actual flu cases. Remarkably, the study unveils that Twitter discussions often anticipate clinical flu encounters by approximately a month, underscoring the predictive potential of social media in tracking disease trends.

This research resonates with the synergy between big data analytics, machine learning techniques like SVM, and intricate spatio-temporal point process models. The fusion of datasets and utilization of sophisticated analytical models notonlyvalidatesthefluoutbreaknowcastingefficacybutalsoexemplifiespotentialinenhancingpublichealth interventions and response strategies.

Transitioning to the second review, Amorós et al. (2020) [75] present a pioneering spatio-temporal Bayesian Markov switching model for swift influenza outbreak detection. This innovative approach strategically incorporates differentiated incidence rates, weaving temporal autoregressive and spatial conditional autoregressive components to capture the intricate spatio-temporal evolution of influenza data. The model's sophistication is evident as it enhances sensitivity, enabling early epidemic peak identification even when initial rates are low. The model's superiority over purely temporal methods materializes through its application to the USA Google Flu Trends database, unveiling improved outbreak detection accuracy and adeptness in handling simultaneous outbreaks originating from distinct time points.

Addressing the pressing need for early influenza outbreak detection, this research elegantly intertwines temporal and spatial methodologies, bolstering surveillance systems. The incorporation of latent variables that classify observations asepidemicorendemicempowersthemodeltoidentifyoutbreakswithvariedonsetsanddurations.Byexpertly capturing the dynamic nature of influenza spread, particularly through the application of spatial structures, the study

Table 5.Outbreak Detection Using Advanced Statistical Methods and Conventional Data Sources

No	Models	Data Source	Result	Limitations/ Future work
1	•Markov Switching Model (MSM) •Seasonal Autoregressive Integrated Moving Average (SARIMA) [54] (2021)	Yearly and monthly data of 49 364 parasitologically-confirmed cases of CLin Isfahan province	The study found significant associations between cutaneous leishmaniasis outbreaks and various environmental factors, such as humidity and wind speed, using both seasonal autoregressive integrated moving average (SARIMA) and Markov switching model (MSM), with the MSM showing improved predictive performance.	Limitations include the omission of important parameters like host-related factors, vectors, parasite type, and health interventions, which are crucial for understanding the epidemiology of cutaneous leishmaniasis in the region and should be considered in future research.
2	•Bayesian structural time series (BSTS) •Autoregressive Integrated Moving Average (ARIMA) [59] (2021)	NIH Islamabad, Pakistan, and Our World in Data: For Iran and India	The BSTS models demonstrated higher forecast accuracy than ARIMA models, contributing to robust predictions. Iran's situation was controlled, while India faced a significant surge.	Limitations: Underreported data due to limited testing, assumptions of current trends continuing, and lack of investigation into risk factors. Future Work: Further study on risk factors, incorporating demographic and social network data, refining forecasting models with more comprehensive testing data.
3	•Latent Markov model [58] (2022)	COVID-19 data from the Italian Civil Protection Department.	The study's analysis of Italian COVID-19 data revealed distinct risk profiles and varying trajectories across regions, capturing the pandemic's dynamics and severity using a Discrete Latent Variable Model.	Limitation: Assumes spatial dependence based on sharing and borders, and potential for biased spatial structure. Future Work: Extend to time-varying latent states, explore more flexible count assumptions, include covariates for interventions, and apply to other disease mapping scenarios.
4	•BSTS •ARIMA [72] (2022)	"Our World in Data" and Humanitarian Data Exchange.	Bayesian structural time series (BSTS) models revealed more accurate COVID-19 predictions	Limitations: Data underreporting, lack of risk factor evaluation, and inherent uncertainty in forecasts

					and assessed the causal impact of vaccinations, indicating successful mortality reduction in the United States and the United Kingdom, while India faces challenges due to slower immunization.	due to data quality. Future Work: Further research on vaccine distribution strategies, incorporating risk factors for improved predictions, and addressing uncertainties in data reporting for enhanced forecasting accuracy.
5	•Bayesian MCMC •extra-Poisson regression	[73]	Taiwanese weekly data	COVID-19	New surveillance metrics were developed to predict domestic cluster infections and guide containment measures against emerging SARS-CoV-2 variants.	Not Available

Table 6.Outbreak Detection Using Advanced Statistical Methods and Social Media or Internet Data Sources

No	Models	Data Source	Result	Limitations/ Future work
1	•SVM •ETAS [74] (2019)	Flu-Related Tweets, CernerHealthFactClinical Encounter Records	Thepotentialoflocation-basedsocialmedia data, integratedwithclinical records and advanced modelingtechniques,toenhance early detection and response to flu outbreaks.	Limitations of the work include: Incomplete geolocation data from social media users. Potential noise and biases in social media data. Aggregation issues due to varying spatial units. Legal, political, and economic obstaclesinutilizing bigdata.Dependence on the quality of data analyzed for accurate insights.
2	•Spatio-temporal BayesianMarkov switching model [75] (2020)	USAGoogleFluTrends database	The proposed spatio-temporal Bayesian Markovswitchingmodel demonstrated superior performance in early detection of influenza outbreaks compared to variationsofthesame model without structured spatial components and a purely temporal model for outbreaks detection.	Not Available
3	•Bayesian hierarchical models •MCMC •BSTM [76] (2021)	Area-levelmobilephone data. Traffic Analysis Zones (TAZs) demographics and mobile phone usercounts.Celltower locations and user counts. Spatial and temporal population distribution data for Shenzhen.	Thestudy effectively utilizedaBayesianspatio-temporal model with area-level mobile phone data to enhance understanding of intra-urban population distribution dynamics.	Limitations: Data Privacy Concerns, SpatialandTemporalResolution Constraints FutureWork:IncorporatingExplanatory Variables, Comparing DifferentTimePeriods,Exploring VaryingEffectsofMixed-UseandUrban Dynamics
4	•Markov Chains [77] (2021)	Twitter data (crawled with keywords: Coronavirus, covid-19, Covid19)		

			proved accuracy.	
5	•HWF •Naïve Bayes classifier •Markov chain •Decisiontree[78] (2021)	From Twitter with keywords related to COVID-19 zones.	The proposed hypergraph-based technique accurately predicted Twitter users' locations based on COVID-19 content, outperforming other algorithms, and can aid in tracking epidemic	Limitations: Reliance on publicly available data. Accuracy linked to local words in tweets. Future Work: Extend the framework to multiple networks. Address geolocation issues in different platforms.

delivers a robust framework for prompt and precise influenza outbreak identification. This framework lends substantial potency to public health interventions.

Next, Wang et al.'s study in 2021 [76] hones in on comprehending population distribution via area-level mobile phone data, employing Bayesian hierarchical models to illuminate space-time patterns. Focused on Shenzhen, China, the research integrates spatial patterns, temporal trends, and deviations through Markov chain Monte Carlo simulation. The model adeptly identifies areas with unstable population trends, thereby informing urban planning and disease response.

By employing Bayesian hierarchical models, the study unravels intricate space-time patterns in intra-urban population distribution. This dual-dimensional approach successfully disentangles predictable trends from unstable fluctuations, enabling the identification of areas with consistent or erratic population changes. The model's incorporation of spatially correlatedanduncorrelatedrandomeffectsaugmentslocalpatterncomprehension.Futuredirectionsencompass exploring alternative spatial priors, integrating explanatory variables' effects, comparing varying temporal periods, and delving into differing coefficients for spatial and temporal influences on population fluctuations. This research contributes substantively to the understanding of urban dynamics and their interplay with space-time patterns, pivotal for efficient urban planning and resource allocation.

The subsequent study, conducted by Suryaningrat et al. (2021) [77], delves into social media's impact in information dissemination, particularly within data mining research. The study investigates the application of Markov Chains to predict the trajectory of COVID-19 discussions on Twitter. Utilizing tweet data collected over distinct observation periods, the research underscores sustained high levels of COVID-19 conversation on Twitter, suggesting its potential as an information hub. The study encourages further refinement by incorporating more comprehensive data and extended observation periods to enhance the Markov Chain model's precision. This enhancement could empower policymakers in utilizing Twitter as a credible COVID-19 information source.

This study delves into the role of social media, notably Twitter, in disseminating and shaping COVID-19 discussions. Leveraging Markov Chains, the research endeavors to predict the trajectory of Twitter discourse surrounding the pandemic. The study encompasses multiple stages, including data collection, cleaning, processing, and the application of the Markov Chain model. The outcomes underscore the sustained nature of COVID-19 discourse on Twitter, indicating the platform's significance as a pandemic information conduit. The research lays a roadmap for future improvement, advocating the integration of more expansive data and longer observation periods to enhance the Markov Chain model's predictive prowess, offering invaluable insights to policymakers relying on Twitter for COVID-19 updates.

The fifth review by Pradeepa (2021) [78] delves into detecting COVID-19's geographic spread through the analysis of Twitter user content. Addressing the challenge of undisclosed user locations, the research introduces a hypergraph-based technique, employing weighting factors termed hypergraph with weighting factor (HWF), to predict users' locations. Byassociatingwordswithlocationsinahypergraph,thismodelboostslocationpredictionaccuracy,thusaiding epidemic zone identification. The technique's superiority over existing methods substantiates its potential for epidemic forecasting and disaster management applications.

This study embarks on analyzing Twitter user content to deduce the geographic spread of COVID-19. The innovative hypergraph-based technique, enhanced by weighting factors, strives to predict users' locations despite the challenge of undisclosed information. Through a multi-step process, the model refines location predictions by mapping words to locations in a hypergraph. The model's efficacy surpasses other methods, spotlighting its potential for forecasting epi-demics and informing disaster management strategies. Despite its effectiveness, the technique remains computationally

feasible for real-time applications, paving the way for a promising avenue of precise location prediction in an evolving epidemic landscape.

4 MACHINE LEARNING TIME SERIES METHODS

Machine learning algorithms have the potential to analyze diverse datasets, encompassing information about known viruses, animal populations, human demographics, biology, biodiversity, physical infrastructures, cultural/social practices worldwide,anddiseasegeolocation,toeffectivelypredictdiseaseoutbreaks.Thissectionisorganizedintothree subsections: section 4.1 focuses on Non-Deep Learning Time Series approaches, section 4.2 reviews Deep Learning Time Series methods, and section 4.3 covers some hybrid time series models. Each part is further categorized into models using conventional datasets and social media data. Tables 7, 10, and 12 provide concise summaries of common methods and their related definitions.

4.1 Non-Deep Learning Time Series Methods

Table 7 highlights some of the most common non-deep learning algorithms used for outbreak detection, utilizing both conventional and social media time series data sources. The purpose of this table is to offer a concise summary of each algorithm, making it easier for readers to comprehend prior works.

*4.1.1 Conventional Data Sources.*This section describes three non-deep learning methods proposed by researchers that focus on conventional data sources. Table 8 represents the proposed methods and their comparison with each other and ground truth, as well as future work and limitations.

In the initial analysis, Chen et al. (2018) [114] delved into the efficacy of employing the machine learning LASSO methodforpredictivemodelingofvariouspathogensacrossdiverseclimaticconditions.Theirapproachinvolved creating distinct LASSO models for each disease, country, and forecast window, thereby assessing the relative significance of predictors under varying contextual settings. The results illuminated distinct cyclical patterns in climatic variables and disease incidence, particularly evident in regions geographically distant from the equator. Furthermore, the study underscoredtheinherentchallengesofachievingaccuratelong-termpredictionswhilehighlightingthesuperior performance of short-term projections, underscoring the crucial need for swift responses to early warning signals.

Turningattentiontothesubsequentinvestigation,thestudybyChenetal.(2019)[115]centeredoninfluenza forecasting utilizing Gaussian process regression techniques. Their innovative approach incorporated meteorological factors to gauge their influence on influenza transmission dynamics. The integration of L1-regularization aided in identifying key explanatory variables contributing to the predictive model. Notably, adjustments were made to the time covariance function to account for non-stationarity and seasonal patterns. Through rigorous comparisons with conventional statistical models, the proposed model showcased its prowess in predicting influenza-like illness (ILI) one week ahead, providing valuable insights into imminent disease trends.

The final review article showcased the contributions of Fang et al. (2022) [93], who devised ARIMA and XGBoost models to predict COVID-19 trends in the USA. A comprehensive assessment of both models' fitting capabilities and prediction accuracies was conducted, resulting in a pivotal discovery. The XGBoost model emerged as a game-changer, significantlyenhancingfitandpredictionaccuracyowingtoitsadeptnessincapturingnon-linearitieswithinthe temporal dynamics of COVID-19 cases.

*4.1.2 Social Media and Internet Based Data Sources.*Public health authorities employ Epidemic Intelligence (EI) to acquire data on disease activity, early warning, and infectious disease outbreaks [116–119]. From a variety of formal

Table 7.Non-Deep Learning Time Series Methods Summary

Model	Interpretation
Least Absolute Shrinkage and Selection Operator (LASSO)[79-83]	LASSO is a statistical formula whose main purpose is the feature selection and regularization of data models. It is based on minimizing Mean Squared Error, which is based on balancing the opposing factors of bias and variance to build the most predictive model. It is usually used in machine learning for the selection of a subset of variables. It can find patterns within large datasets while avoiding the problem of over-fitting.
Least-Angle Regression (LARS)[84,85]	LARS is a variable selection method with proven performance for cross-sectional data. It is used in regression for high-dimensional data (i.e., data with a large number of attributes). it is extended to time series forecasting with many predictors.
XG-Boost[1,86-94]	XGBoost is an implementation of the gradient boosting ensemble algorithm for classification and regression. XGBoost uses the ensemble of weak prediction models, gradient boosting helps us in making predictions. Examples of weak models can be decision trees. It helps in generalizing the other model by optimizing the arbitrary differential loss function. Time series datasets can be transformed into supervised learning using a sliding-window representation for XG-Boost.
SupportVectorMachine(SVM) [95-97]	SVM, a supervised machine learning algorithm, classifies and analyzes data for classification andregression.Itcategorizesdata,creatingwidemarginsbetween categories.SVMfinds applications in text, images, and time series forecasting. In time series, it maps data to a higher dimension for separation, and for regression, it forecasts nonlinear, non-stationary data with undefined processes.
DecisionTrees[98-103]	Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously splitaccordingtoacertainparameter.Thetreecanbeexplainedbytwoentities,namely decision nodes, and leaves.
GaussianNaïveBayes(GNB) [104-108]	Naïve Bayes is a generative model. (Gaussian) Naïve Bayes assumes that each class follows aGaussiandistribution.ThedifferencebetweenQDAand(Gaussia n)NaïveBayesisthat Naïve Bayes assumes independence of the features, which means the covariance matrices are diagonal matrices.
RandomForest[109-113]	Random Forest is a supervised machine learning algorithm made up of decision trees. Random Forest is used for both classification and regression—for example, classifying whether an email is “spam” or “not spam”.

and increasingly informal sources, EI systems routinely compile official reports and rumors of probable outbreaks [120].

To detect information on disease outbreaks, tools like the Global Public Health Intelligence Network (GPHIN) and Medisys collect data from international media sources like news wires and websites [116]. Google’s Flu Trends research, which evaluated flu activity by compiling real-time web search queries for flu-related terms [121], has shown how these systems might be improved. The information kept in commercial search query logs, which may be linked to EI systems, has the issue of not being publicly accessible. However, the rise of user-generated content on social networking sites like Facebook and Twitter gives EI systems a very easy way to access data on current online activity. There are already more than 15 million unique users per month using Twitter [12], a micro-blogging site that enables users to publish and read 140-character messages, or "tweets," from other users [122]. Twitter gives third parties the ability to search user

messages and retrieve the text along with user-specific data, such as the poster's location, in a format that is simple to

store and analyze. Table 9 depicts the proposed methods and their comparison with others and ground truth, as well as

future work and limitations.

Table 8.Outbreak Detection Using Non-deep Learning Methods and Conventional Data Sources

No	Models	Data Source	Result	Limitations/Future work
1	•LASSO regression [114] (2018)	National Institute of Infectious Diseases (NIID)- Japan, Bureau of Epidemiology, Ministry of Public Health-Thailand, Ministry of Health- Singapore, Taiwan National Infectious Disease Statistics System -Taiwan, Weather Underground - Taiwan, Thailand, Singapore, Japan Meteorological Agency - Japan	Regression models utilizing LASSO can notably enhance predictive accuracy for certain diseases using a specific set of variables; however, for other diseases, simpler models yield comparable outcomes to more intricate ones. Regular implementation of models based on this approach reveals superior short-term disease predictions compared to long-term forecasts, implying the necessity for rapid response capabilities in public health agencies to address early indications of potential infectious disease outbreaks.	limitations include its divergence from traditional epidemic models like compartmental and network models, as well as its reliance on capital city weather data, overlooking regional climatic variations; future accuracy may increase with finer data resolution.
2	•Gaussian process regression+LASSO •Linear regression •ANN •SVR •SARIMA [115] (2019)	Influenza-like illness (ILI) Shenzhen (CDC)	Gaussian process regression + LASSO model outperforms other models in terms of one-week-ahead prediction of influenza-like illness.	As a future work, the authors intend to examine the efficacy of the proposed method in addition cities within the next few years. In addition, spatial information will be incorporated to model the effect of meteorology on influenza's spatial-temporal spread.
3	•ARIMA •XGBoost [93] (2022)	USA COVID-19 cases and vaccination From CDC	Based on the daily case numbers from the previous 7 days, the fit and prediction accuracy for the following 14 days are significantly improved by the XGBoost model.	This work has the following limitations: first, the study period was relatively short and should have been expanded to better reflect the future development of COVID-19 in the United States. And secondly, the XGBoost model was created using pre vaccination-induced herd immunity as its foundation. Consequently, as the incidence of transmissible variants rises, the accuracy of predictions may diminish.

The initial study under review, conducted by Jain and Kumar in 2015 [123], devised a method for tracking the flu

epidemic in India during February to March 2015, employing Twitter content. An innovative approach based on dynamic

keywords sourced from RSS feeds was introduced to retrieve tweets via Twitter. Data spanning 60 days, commencing

from February 1, 2015, and concluding on March 31, 2015, was collected from Twitter. This dataset was instrumental

in monitoring key terms associated with "H1N1" or "swine flu" over time. The authors conducted content analysis of

tweets and scrutinized states most significantly impacted by the rapidly spreading flu.

Employing sentiment analysis and a count-based technique, the dataset was dissected to unravel pivotal aspects during

the Influenza-A (H1N1) pandemic. The study systematically explored various parameters to extract pertinent information

about the disease and gauge the general awareness surrounding it. Classification was deployed to differentiate real-time

Table 9.Outbreak Detection Using Non-deep Learning Methods and Social Media or Internet Data Sources

No	Models	Data Source	Result	Limitations/ Future work
1	•SVM •Naïve Bayes •Random Forest •Decision Tree [123] (2015)	Keywords from tweets and RSS feeds Keywords from forH1N1orSwineflu in India	The analysis of data sets and classi-fication results in the development ofanearlywarnings ystemthat detectsanimpendi ngspikeinan epidemic before the official surveil-lance systems are examined. SVM performed best in classification as well.	Not Available
2	•Decision Tree •Random Fores •SVM •Naïve Bayes •AdaBoost •KNN •FastText [124] (2019)	Twitterinfluenzasur veil-lance dataset	In predicting flu outbreaks, the com-bination of FastText (FT) classifica-tionwitharegressionmo delout-performs other classification algo-rithms.	Not Available
3	•Decision trees •Random forests •SVM •SVM-Perf [125] (2020)	Twitter posts between 2011 and 2014 in key cities oftheaffectedregionin West Africa	The results show that the adapted architectureu singSVM-Perfob-tains more relevant alerts than the others.	The proposed architecture could be adapted for other social media plat-forms or disease types in the future.
4	•SVM •NaïveBayes[126] (2020)	Twitterusingkeywords, language,Geolocationre-lated to the Covid-19	TheNaïveBayesclassifier per-formed better than SVM at identify-ing tweets related to Covid-19.	Not Available
5	•RF •KNN •SVM •DT [127] (2021)	Tweets on dengue and flu from Twitter	Intermsoffindingtweetsr elated toseasonaloutbreaks,th eresults showedthattheRFclassifierout-performed SVM, DT, and KNN.	Therearesomedrawbackstothe proposedmodel.Becausesuper-visedlearningwasusedinthis study, the data used for model train-ing had to be labeled. To avoid the needforlabeleddata,themodel should be trained unsupervised. In thefuture,theproposedmodel could be used as a surveillance sys-temtodetectthespreadofcoron-avirus and COVID-19.

data from noise or irrelevant tweets. Four distinct algorithms (SVM, Naïve Bayes, Random Forest, and Decision Tree)

were employed for classification, with the SVM classifier demonstrating superior performance.

This investigation highlighted the potential of using social media to comprehend public health dynamics, illustrating

Twitter's utility in detecting disease outbreaks by analyzing data generated within the realm of social media.

Shifting the focus to the work by Alessa and Faezipour in 2019 [124], the authors formulated a comprehensive framework for predicting outbreaks in 2019. Comprising three key modules—text classification, mapping, and linear

regression—the framework was designed to forecast weekly flu rates. The text classification module harnessed sentiment analysisandpredefinedkeywordoccurrences.Twitterinfluenzasurveillancedatasetsweregatheredandvarious classifiers, including FastText (FT) and six conventional machine learning algorithms, were evaluated to ascertain optimal effectiveness for the framework. The classified Twitter documents and historical CDC data were then fed into a linear regression-based module for predicting weekly flu rates. Remarkably, the proposed FastText (FT) classification emerged as the most efficient and accurate model. Impressively, the final flu trend prediction based on Twitter documents exhibited a robust Pearson correlation of 96.29 percent with the CDC's actual data from the initial months of 2018.

Moving on to the research by Joshi et al. in 2020 [125], two variations of an existing surveillance architecture were examined. The first version aggregated tweets related to various symptoms, while the second version considered each symptom separately before amalgamating the set of alerts produced by the architecture. This study utilized a database of tweets from impacted areas in West Africa between 2011 and 2014, focusing on the Ebola symptoms of fever and rash. The results led to two crucial conclusions: social media provided an early warning three months prior to the 2014 Ebola pandemic, and data aggregation could potentially yield more frequent notifications than alert aggregation. The SVM and SVM-Perf classifiers were employed for personal health mention categorization. The findings indicated that the modified architecture utilizing SVM-Perf produced more relevant alerts compared to the SVM-only architecture.

In the study conducted by Fakhry and colleagues in 2020 [126], a dual machine learning approach was employed to analyze present and future Covid-19 case data using publicly available social media information. Through a combination of machine learning and classical data mining techniques, disease cases were estimated based on social media content inaspecificgeographicregion.Thesentimentanalysiswasleveragedtogaugethepublic'sperceptionofdisease awareness in the same location. Employing specific keywords, tweets related to Coronavirus were extracted from Twitter and classified using Naïve Bayes and SVM classifiers. Strong correlation was observed between classified tweets and real-world data.

Lastly, Amin et al.'s algorithm in 2021 [127] aimed to detect seasonal outbreaks using Twitter data through machine learningapproaches.Twocategoriesoftweets—disease-positiveanddisease-negative—wereemployedtoidentify outbreaks related to dengue and flu. Machine learning algorithms including Random Forest (RF), K-Nearest Neighbor (KNN),SupportVectorMachine(SVM),andDecisionTree(DT)wereharnessed,alongwithTermFrequencyand Inverse Document Frequency (TF-IDF) for feature extraction. Notably, the RF classifier outperformed others in terms of accuracy, precision, recall, and F1 measure. Despite its findings, the model employed supervised learning, highlighting a need for further exploration of unsupervised training approaches.

4.2 Deep Learning Time Series Methods

Deeplearninghasbeenappliedinhealthcarethroughmedicalimaging,chatbotsforpatterndetectioninpatient complaints, algorithms for cancer detection, and systems identifying rare diseases or pathology. It provides valuable insights to medical professionals, enabling early problem detection and more personalized care. Table 10 highlights common deep learning methods used for outbreak detection with conventional and internet time series data.

*4.2.1 ConventionalDataSources.*Deeplearninginhealthcarehasreceivedalotofattentioninrecentyearsasa crucial tool for assisting in clinical decision-making and disease diagnosis [152, 153]. In late December 2019, a Canadian company (Blue Dot) correctly notified the location of Covid-19 outbreak, demonstrating the effectiveness of AI and deep learning in the current pandemic for outbreak prediction. The development of image verification to differentiate COVID-19 pneumonia from other benign respiratory illnesses also benefited from AI [154]. We provide a summary

Table 10.Deep Learning Time Series Methods Summary

Model	Interpretation
LSTM[128–132]	LSTM stands for Long short-term memory. LSTM cells are used in recurrent neural networks thatlearntopredictthefuturefromsequencesofvariablelengths.Themainideabehind LSTMcellsistolearntheimportantpartsofthesequenceseensofarandforgettheless important ones. An LSTM network is a recurrent neural network (RNN) that processes input data by looping over time steps and updating the network state. The network state contains information remembered over all previous time steps. LSTM network can be used to forecast subsequent values of a time series or sequence using previous time steps as input.
Bi-LSTM[133–135]	Bidirectional LSTM, or biLSTM, is a sequence processing model that consists of two LSTMs: one taking the input in a forward direction, and the other in a backward direction. In problems where all timesteps of the input sequence are available, Bidirectional LSTMs train two instead of one LSTMs on the input sequence.
TBAT[136–139]	TBATS model is a forecasting model based on exponential smoothing. The name is an acronym for Trigonometric, Box-Cox transform, ARMA errors, Trend, and Seasonal components. The TBATS model's main feature is its capability to deal with multiple seasonalities by modeling each seasonality with a trigonometric representation based on the Fourier series.
N-BEATS[137,139,140]	N-BEATS is a deep neural structure featuring backward and forward residual links and a deep stack of fully connected layers. It takes an entire historical data window and generates multiple forecasttimepointssimultaneouslythroughextensiveuseoffullyconnectedlayers.The architecture comprises interconnected blocks in a residual manner: the initial block models pastandfuturedata,whilesubsequentblocksfocusonresidualerrorsfromtheprevious reconstruction, updating forecasts accordingly. This residual-based design enables a deep stack of blocks without gradient vanishing concerns and offers benefits akin to boosting/ensembling techniques, where predictions from various blocks are combined, with each block capturing different aspects of the forecast.
FBProphet[141–145]	FB Prophet is a procedure for forecasting time series data based on an additive regression model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. it is great for stationary data. Stationary data is time series data that follow similar behavior and have the same statistical properties throughout time.
DeepAR[136–139,146–148]	DeepAR, a supervised time series forecasting algorithm, employs recurrent neural networks (RNN) to generate both point and probabilistic predictions. This method considers not only past values but also incorporates additional covariates like dynamic historical features, static attributes,andfutureevents.However,traditionaltechniquesoftentreateachtimeseries independently, missing out on potential cross-learning opportunities and valuable information relevant to the specific use case.
NARNN[149–151]	The Nonlinear Autoregression Neural Network (NARNN) model is a technique that performs nonlinear regression through the neural network.

of eleven recent studies that used deep learning to anticipate and identify outbreaks in this section. The methods, data sources, and findings used by earlier researchers to show and compare the effectiveness of their methods are summarised in Table 11.

The ongoing COVID-19 pandemic has triggered substantial societal disruptions, compelling governments to implement drastic measures to curb its spread. Anticipating the outbreak's peak could significantly mitigate its impact, enabling governments to tailor policies, plan preventive actions, enhance public health communication, and bolster healthcare systems. As outlined in Table 11, the initial review by Papastefanopoulos et al. in 2020 [139] examined the

Table 11.Outbreak Detection Using Deep Learning Methods and Conventional Data Sources

No	Models	Data Source	Result	Limitations/Future work
1	•ARIMA •Holt–Winters additive •TBAT •Facebook's Prophet •DeepAR •N-Beats [139] (2020)	NovelCoronaVirus2019dataset And population-by-country dataset	TheARIMA,Holt–Winters,and TBAToutperformthedeeplearning models due to the small dataset size.	Future works include enhancing prediction accuracy, potentially improved? by adopting ensembles to minimize errors, multivariate time series modeling considering various pandemic-related factors, and transferring knowledge across countries.
2	•DeCoVNet [155] (2020)	Data from the National Health Commission of the People's Republic of China Diagnosed with COVID-19	DeCoVNet shows promise for precise and swift COVID-19 diagnosis, aidingfrontlinemedicalstaffand global epidemic control.	Limitations include Network design and training that could improve? data from a single hospital lacks cross-validation, and the algorithm's COVID-19 diagnosis is hindered by deep learning's early-stage explainability.
3	•ARIMA •EHW •Linear Regression •SVM Regression •FDTL •MLP Regression •PNN •PNN+cf [32] (2020)	2019-nCov from Chinese health authorities	Thefindingsdemonstratethat PNN+cfiseffectiveinproducing reliable forecasts during the crucial periodofdiseaseoutbreakswhen the samples are scarce.	Future work should analyze variations in forecasting outcomes among algorithms, test more panel selection methods, explore prominent algorithms designs, and expand the methodology.
4	•Stacked LSTM •Bi-LSTM •Conv-LSTM [135] (2020)	Covid-19 data from the Ministry of Health and Family Welfare (Government of India)	Bi-LSTM is better suited for prediction purposes of 1 to 7 days in the future.	Not Available
5	•Stacked LSTM •Bi-LSTM •Conv-LSTM [134] (2020)	Indian data is sourced from the Ministry of Health and Family Welfare, andUSdatafromtheCentersfor Disease Control and Prevention.	ConvLSTMmodeloutperformed theothertwomodelsforpredicting the next month's case numbers.	Future research: Analyzing COVID-19's economic impact across sectors and devising recovery strategies; investigating aerosol transmission and global case prediction.
6	•ARIMA •SVR •LSTM •Bi-LSTM •GRU [156] (2020)	Covid-19 dataset for different countries.	Bi-LSTM outperforms others in predictingCovid-19.Dataforeach country includes given cases for 110 days and must be predicted for the next 48 days.	Not Available
7	•ARIMA •NARNN •LSTM [151] (2020)	Covid-19 data from European Center for Disease Prevention and Control for 8 different European countries (Denmark, Belgium, Germany, France,UnitedKingdom,Finland, Switzerland, and Turkey)	Basedontheresults,itwasdeterminedthattheLSTMapproachis significantly more successful than ARIMA and NARNN to predict the number of cases for the next 7 days.	Not Available
8	•Bi-LSTM •ARIMA •(SMA-6), Double Exponential •D-EXP-MA [157] (2021)	Covid-19 case numbers for different countries from their health authorities' websites	Bi-LSTM outperforms other models using 6 previous days' cases to predict the next day's case.	Not Available
9	•LSTM •XGBoost [158] (2021)	USACOVID-19caseswerecollectedfromtheWorldHealthOrganization website.	LSTM outperforms XGBoost in predictingthenextdaycasenumber based on the previous 7 days of data.	Future research: Modeling lockdown-easing scenarios and their impact on case trends; assessing vaccination effects on COVID-19 outbreak in Qatar.
10	•LSTM •GRU •CNN [159] (2022)	COVID-19 datasets for 10 countries from Humanitarian Data Exchange (HDX).	The results showed that after data augmentation, the performance of the LSTM and CNN models significantly improved. Furthermore, the proposed method achieves an average performance for GRU.	Limitations include lack of re-intervention/vaccination data and sole liance on infection time series. Future: Investigate time series augmentation methods like dynamic time-warping barycentric averaging.
11	•LSTM •CDC ensemble model [160] (2022)	CDC and CSSE Johns Hopkins University data related to infectious diseases in the United States	Thestudydemonstratedthatthe LSTM model outperformed existing methods, including the CDC ensemblemodel,inforecastingCOVID-19casesanddeaths,highlighting theeffectivenessofincorporating diverse data sources and advanced deep learning techniques.	One limitation of this study is the reliance on self-reported data, which may introduce bias

accuracy of diverse time series modeling techniques for identifying coronavirus epidemics across ten countries with the highest confirmed cases by May 4, 2020. Six time series methodologies—ARIMA, Holt-Winters additive model (HWAAS), TBAT, Facebook's Prophet, DeepAR, and N-Beats—were developed and compared for each country. Publicly available

datasets on virus progression and population size were employed for model development and analysis, acquired from "Novel Corona Virus 2019 Dataset" [161] and "population-by-country dataset" [162] on kaggle.com.

Results showcased that while no single approach was universally optimal, conventional statistical methods like ARIMAandTBAToutperformeddeeplearningcounterpartssuchasDeepARandN-BEATS,whichalignedwith expectations given limited data availability. Specifically, ARIMA and TBAT exhibited superior performance across seven of ten cases. Statistical analysis employing Friedman's test and Holm's post-hoc analysis validated the superiority of TBAT over Prophet, DeepAR (Gluonts), and N-BEATS.

The COVID-19 outbreak placed significant strain on Hubei province, leading to a substantial volume of chest CT scans for suspected patients. The shortage of medical professionals exacerbated the risk of missed diagnoses for minor lesions.

The subsequent review, conducted by Zheng et al. in 2020 [155], introduced a 3D deep convolutional neural network (DeCoVNet) for COVID-19 detection through 3D CT volumes. The model encompassed lung segmentation via a pre-trained UNet, followed by a 3D deep neural network to predict COVID-19 infection likelihood. A dataset of 499 CT volumes for training and 131 volumes for testing were used. The algorithm achieved high ROC AUC, PR AUC, accuracy, sensitivity, specificity, and predictive values for COVID-positive and COVID-negative classification.

Fong et al. in 2020 [32] addressed the challenge of forecasting at the initial stages of an epidemic with scarce data. Their approach, Group of Optimized and Multi-Source Selection (GROOMS), combined multiple forecasting models, includingpolynomialneuralnetworks(PNN),forgrouppredictions.ExperimentsshowedthatPNN,particularly PNN+cf, excelled in generating accurate forecasts even with limited data.

Aroraetal.in2020[135]employeddeeplearningmodels,specificallyrecurrentneuralnetwork(RNN)-based LSTM variations, to forecast COVID-19 cases in Indian states. Accurate short-term predictions were achieved using Bi-directional LSTM.

Shastri et al. [134] employed LSTM, Stacked LSTM, and Convolutional LSTM to comparatively analyze COVID-19 cases in India and the USA. Convolutional LSTM outperformed the other models, providing accurate predictions and highlighting the significance of various factors beyond model choice.

Shahid et al. in 2020 [156] presented forecast models for COVID-19 cases in ten countries, utilizing ARIMA, SVR, LSTM, Bi-LSTM, and GRU models. Bi-LSTM consistently demonstrated superior performance, offering robust and accurate predictions for pandemic planning.

Kirbas et al. in 2020 [151] modeled COVID-19 cases using ARIMA, NARNN, and LSTM techniques. LSTM emerged as the most precise model for short-term prediction, providing insights into the outbreak's trajectory.

Saidetal.in2021[157]introducedadeeplearningapproachforpredictingdailycumulativeCOVID-19cases, grouping countries with similar characteristics. Their method significantly improved prediction performance compared to existing techniques.

Luo et al. in 2021 [158] harnessed LSTM and XGBoost algorithms for predicting daily confirmed cases in the US. Their analysis emphasized the importance of precautionary measures and accurate data for meaningful predictions.

Abbasimehr et al. in 2022 [159] employed time series augmentation techniques to enhance deep learning model accuracy. The proposed augmentation approach exhibited superior performance across multiple countries, improving forecasting accuracy.

Lastly, Du et al. in 2022 [160] presented a novel multi-stage deep learning model for forecasting COVID-19 cases and deaths in the United States. The model, referred to as the LSTM model, utilizes a comprehensive dataset encompassing epidemiological, mobility, survey, climate, demographic, and genomic data. Through rigorous evaluation, the LSTM

Table 12.Outbreak Detection Using Deep Learning Methods and Social Media or Internet Data Sources

No	Models	Data Source	Result	Limitations/ Future work
1	•DNN •LSTM •ARIMA •OLS [163] (2018)	KCDC,searchquerydata fromSouth Korean-specific search engines, Twitter social media big data, and weather data such as temperature and humidity	Deep learning models DNN and LSTM predicted infectiousdiseaseswith a7-daylagsignificantly better than OLS and ARIMA models.	Limitationsincludethreefactors: the study's brief data collection period, regionally combined predictions, and consideration of a constrained set of deeplearningmodelparameters.
2	•SSL •SVM •DNN [164] (2020)	Articles and reports provided by Medisys related to disease	SVM,whichconsistently performs well across fields;SSL,whichperformswellwhenlabel imbalanceddatasetsare used; and DNN, a trending methodwithoutstanding performance, were used to predict disease occurrence. data could enhance future predictions.	Limitationsincluderequiring systematicdatatimeframeselection, especially for seasonal diseases. Initial data limitations andmodelperformancesuggest potential for improvement. Exploringsimilardiseasepatternsinothernationsandintegrating global air passenger

modelconsistentlyoutperformstheCDCensemblemodelforallevaluationmetrics,particularlyinlonger-term forecasting. The study emphasizes the importance of considering various factors when forecasting COVID-19 outbreaks. These factors include outbreak phase, location, time, and the availability of genomic data. The authors highlight the need for careful model selection and evaluation to ensure accurate and reliable predictions. Additionally, the study demonstrates the value of incorporating genomic surveillance data to enhance forecasting accuracy, especially during periods of emerging variants.

*4.2.2 Social Media and Internet Based Data Sources.*Social media allows users to share news, ideas, and opinions globally, and researchers can automate the collection and analysis of posts for more effective disease surveillance [11]. Social media and internet search data have been successfully used to monitor outbreaks like Zika, Dengue, MERS, Ebola, and COVID-19 [10]. This section reviews various outbreak detection methods, evaluation techniques, and challenges in using social media and internet-based data. It also presents findings on two deep learning methods, their data sources, results, and limitations, as shown in Table 12.

The first review discusses the work by Chae et al. in 2018 [163], where the researchers aimed to predict infectious diseases using deep learning algorithms by incorporating various sources of data, including infectious disease occurrence data from the Korea Center for Disease Control (KCDC), search query data from South Korean-specific search engines, Twitter social media data, and weather data like temperature and humidity. The search queries used included both thedisease'snameandits symptoms.The studydevelopedaninfectious diseasemonitoringmodelby combining non-clinical search data, Twitter data,

and weather data. The researchers created various Ordinary Least Squares (OLS) models with different combinations of variables and evaluated their explanatory power using corrected R-squared values. They introduced a lag parameter of 1-14 days for infectious diseases and selected a seven-day lag as the best parameter for prediction. They built OLS, ARIMA, Deep Neural Network (DNN), and Long Short-Term Memory (LSTM) models using the chosen parameters. Comparing these models, they found that DNN and LSTM models outperformed

Table 13.Hybrid Time Series Methods Summary

Model	Interpretation
VAR-LSTM[166]	In this method the data are first trained by using the Vector Auto-regressive (VAR) technique, then the outputs are used as the inputs for the Long Short-Term Memory (LSTM) networks by using a deep learning (DL) approach.
CNN-LSTM[167–172]	In this method LSTM models can easily capture sequence pattern information, but they are tailored to deal with temporal correlations and only use the features specified in the training set. Another popular deep learning method is convolutional neural networks (CNNs). CNN models are capable of filtering out noise in the input data and extracting more valuable knowledge for the final forecasting model.
SVM-RBM[173,174]	Restricted Boltzmann Machines are stochastic two-layered neural networks that belong to a category of energy-based models that can detect inherent patterns automatically in the data by reconstructing input. They have two layers visible and hidden. RBM used to serve as features auto encoding processor for SVM.

OLS and ARIMA models. DNN was more consistent when diseases were spreading, while LSTM was more accurate. The authors suggested that their models could help reduce reporting delays in existing surveillance systems, leading to cost savings.

The second review pertains to the research by Kim and Ahn in 2021 [165]. They created three machine learning models to detect early outbreaks of infectious diseases using disease-related media articles and reports. The models used were Support Vector Machine (SVM), Semi-Supervised Learning (SSL), and Deep Neural Network (DNN). The authors collected Medisys data from January to December 2019, including article titles, descriptions, published dates and times, disease information, and geographic coordinates. The daily article counts related to each disease by country were structured as a numerical dataset. The models were trained on data from one three-month period and tested on the subsequent three-month period for prediction. SSL showed the best performance, followed by SVM and DNN. All three models achieved average accuracies greater than 0.7 and F1 scores greater than 0.75. The proposed models were seen as useful for preparing for future infectious disease outbreaks, particularly in countries with limited disease surveillance systems.

4.3 Hybrid Time Series Methods

Methods or approaches that are made up of two or three methods or algorithms are referred to as hybrid methods. This section elaborates on some of the strategies presented by authors for outbreak detection utilizing conventional and internet-based data. Table 13 depicts some of the hybrid models used by authors to detect outbreaks from conventional and social media/internet time series data.

*4.3.1 Conventional Data Sources.*The five most recent and innovative hybrid methods for outbreak detection are the focus of this section. Table 14 includes a column for future work and limitations that helps researchers come up with some ideas for their work while also summarising these methods' names, data sources, and results.

The first review discusses the work by Pustokhin et al. in 2020 [176], where they introduced the RCAL-BiLSTM model for COVID-19 diagnosis. The model involves preprocessing images using bilateral filtering (BF) to remove noise, followed by feature extraction using RCAL-BiLSTM, and classification using a softmax (SM) layer. The RCAL-BiLSTM model includes ResNet-based feature extraction, Class Attention Layer (CAL), and Bidirectional LSTM modules. The

Table 14.Outbreak Detection Using Hybrid Methods and Surveillance Data

No	Models	Data Source	Result	Limitations/ Future work
1	•RCAL-BiLSTM •CNN,DTL,MLP,LR,XGBoost,KNN,DT,Xiaowei Xu et al. [175] (2020)	Chest-X-Ray dataset [176]	RCAL-BiLSTM model outperforms othermodelsandcanbeincorporated in real-time hospitals to predict and classify the COVID-19 pandemic.	Not Available
2	•SEIR •LSTM [164] (2020)	Domesticmigration data and Covid-19dataof China	ThedynamicSEIRmodel,combinedwithLSTM-basedAI,effectivelypredicteddCOVID-19epidemic peaks and sizes.	Limitations include not considering variableslikediagnosticcapacity that could impact case numbers and notaccountingfor seasonalinfluences.
3	•VAR-LSTM •LSTM [166] (2020)	Johns Hopkins University and the Canadian Health Authority	ThepositiveCovid-19casenumbers can be predicted by VAR-LSTM with high accuracy. It employed lag 2 data to forecast the case numbers for the upcoming 15 days, outperforming the LSTM in the process.	Not Available
4	•Original Ensemble forecasts,RoundedEnsemble forecasts [177] (2020)	CenterforSystems Science andEngineering (CSSE)atJohns HopkinsUniversity	Accurateshort-termpredictions withwell-calibratedintervals(92-96% coverage), though accuracy declines for longer forecasts.	Futureworkinvolvesenhancing model|calibrationandexpanding the ensemble to improve long-term forecast accuracy and robustness.
5	•MLP •ANFIS •SIR/SEIR [178] (2020)	Data collected fromworldometers for five countries, including Italy, Germany, Iran, USA, and China	Theresearchdemonstratestheeffectiveness of hybrid ML-statistical models in COVID-19 outbreak prediction.	Futureresearchshoulddevelop country-specific ML models to address the unique characteristics of COVID-19outbreaks.Giventhe variability between outbreaks, creatingasingleglobalmodelwith broadgeneralizationmaybechallenging.
6	•CNN-LSTM •Other baselines (CNN, LSTM, ARIMA, FBProphet, LR, Ridge, Lasso, XGBoostR, AdaBoostR, RFR, GBR, ETR,	WHO COVID-19 dashboard for all countries	When compared to 17 baseline time series forecasting models, the proposed CNN-LSTM model will outperform them all.	Futureworkinvolvesenhancing COVID-19forecastingaccuracy byincorporatingadditionaldata andexternalfactorslikeseasonal changes,vaccinationplans,and

	BaggingR, GPR, SVR, DTR, KNNR) [169] (2021)			lockdowns.Exploringvariousre-samplingandrestructuringmethods is planned. Additionally, developing an uncertainty management strategytoquantifyandconvey pandemicinformationmoreeffectively is essential.
7	•LR ,Polynomial regression , LSTM , GRU , RNN , ARIMA , Prophet •LSTM-GR [179] (2022)	COVID-19cases in India from Kaggle for each state	Thestudyfindingsrevealthat theproposedstackedLSTM-GRU modeloutperformsalloth ermodels.	Future work could incorporate new components and algorithms into the hybrid model to address the prevalence of asymptomatic cases in India and improve result accuracy.
8	•MLR , LSTM , Prophet , SEIR •XGBoost-LSTM(XLM) [180] (2022)	Highly pathogenic infectious disease transmission dataset published by Baidu	Theexperimentsshowthatthe XLMpredictionframeworkproposedinthispaperoutperforms other prediction methods.	Future work includes enhancing the framework using big traffic data's distinct features to improve disease predictionaccuracy,emphasizing key transmission attributes and expanding its applicability to support disease prevention.

model's performance was evaluated on a Chest X-ray dataset, and it outperformed other models, achieving a higher F-score value of 93.10

The second review by Yang et al. in 2020 [164] integrated population migration data with a susceptible-exposed-infected-removed(SEIR)modelandusedanAI-basedLSTMapproach,trainedon2003SARSdata,topredictthe progression of COVID-19. Their dynamic SEIR model effectively captured the epidemic peaks and trends, highlighting the crucial role of control measures implemented on January 23, 2020, in reducing the eventual size of the epidemic.

The third review pertains to research by Afzali et al. in 2020 [166], where they introduced a hybrid VAR-LSTM model for COVID-19 data modeling. The model combined Vector Auto-regressive (VAR) and Long Short-Term Memory (LSTM) techniques. The VAR approach was used to train the LSTM network, improving its ability to forecast confirmed COVID-19cases.ThehybridVAR-LSTMmodeloutperformedLSTMintermsofaccuracyandperformancewhen compared with actual case data.

In the fourth work by Evan L. Ray et al. in 2020 [177], they analyzed the real-time application of an open, collaborative ensemble forecasting model to predict COVID-19-related deaths in the United States. Their study, spanning from April toJuly2020,aggregatedprobabilisticforecastsfrommultiplemodelstocreateaweeklyensembleforecast.Each participating model provided predictions for one to four weeks ahead, including median estimates and a range of prediction intervals to account for uncertainty. The ensemble was constructed by averaging these predictions, providing a more robust and reliable forecast. The results indicated that these ensemble forecasts provided accurate short-term predictions, particularly within a one-week horizon, though the accuracy diminished at longer horizons. The ensemble's prediction intervals were well-calibrated, with observed outcomes falling within the predicted ranges, demonstrating the model's reliability for public health decision-making during the pandemic.

The fifth review focuses on the research proposed by Ardabili et al. in 2020 [178] that investigated the potential of combining machine learning (ML) with statistical models for predicting COVID-19 outbreaks. It compares ML algorithms like Multi-Layer Perceptron (MLP) with 8, 12, and 16 neurons, and Adaptive Network-Based Fuzzy Inference System (ANFIS) with Triangular, Trapezoidal, and Gaussian membership functions, with traditional epidemiological models, finding that ML models offer superior accuracy and generalization. The Grey Wolf Optimizer (GWO) is identified as the best parameter tuning method for statistical models. The logistic model consistently outperforms other statistical models.BothweeklyanddailydatasamplingareeffectiveforMLmodeling.ThestudysuggestsMLcanaddress challenges in traditional models, such as missing data. Integrating ML with SIR/SEIR models is proposed for enhanced accuracy and predictive power. Overall, the research demonstrates the effectiveness of hybrid ML-statistical models in COVID-19 outbreak prediction, offering a promising alternative to traditional approaches.

The sixth work was done on a time-series dataset by Zain et al. in 2021 [169]. The researchers developed a hybrid CNN-LSTM model to predict the number of confirmed COVID-19 cases. They used WHO COVID-19 dashboard data and compared their model with 17 baseline models. The CNN-LSTM model showed superior performance, indicating that combining both models in an encoder-decoder structure significantly improved forecasting accuracy. The CNN-LSTM model effectively handled noise through convolutional layers and captured short- and long-term relationships in the time series.

The seventh review discusses the work by Sah et al. in 2022 [179], where they proposed a hybrid stacked LSTM-GRU model to predict COVID-19 cases in India. The stacked LSTM output was used as input for the GRU model, resulting in better prediction accuracy compared to other state-of-the-art models.

The last review discusses the work by Guo et al. in 2022 [180], where a mixed XGBoost-LSTM (XLM) framework was introduced to predict the spread of infectious diseases across multiple cities and regions. The framework utilized Baidu's

infectious disease transmission dataset and combined K-means clustering, XGBoost modeling, and LSTM modeling. The XLM framework demonstrated better predictive performance compared to other methods for predicting infectious disease spread.

These reviews highlight various innovative approaches to utilizing deep learning models for COVID-19 diagnosis, case prediction, and transmission forecasting, each contributing valuable insights to the field of disease prediction and control.

4.3.2 Social Media and Internet Based Data Sources. Health authorities maintain a surveillance system to limit the spread of infectious diseases, but missing and delayed information makes timely action challenging. Predicting illness patterns is also difficult due to their unpredictability. To address these issues, a data-driven infectious disease prediction model is needed, which could help reduce societal costs by forecasting trends. Recognizing this, researchers are increasingly focusing on data-driven studies to enhance current systems and develop new models [181–187]. Many of these studies use large datasets like Internet search queries [13, 28, 188–191], which can be processed in near-real-time. Social media big data is also being considered. Table 15 presents recent studies utilizing hybrid models with social media data for outbreak detection.

The first reviewed article by Xiao et al. [173] in 2016 demonstrates how social media data can be harnessed using data mining techniques to detect real-world phenomena. They treat microblog users as "sensors," using flu-related group posts as early indicators. By collecting microblog posts from Sina Weibo, they create a crowdsourced data approach. They propose supervised and unsupervised methods to estimate flu-infected individuals, focusing on sentiment scores inanunsupervisedmodel.TheyenhanceasupervisedmodelwithConditionalRandomFields(CRFs)forquicker flu outbreak detection. Their hybrid classification model using a Support Vector Machine and Restricted Boltzmann Machine (SVM-RBM) classifies flu-related posts, enabling real-time flu detection. Both models outperform others when considering sentiment analysis and additional features. The proposed method is effective, as demonstrated by testing on real-time datasets for various applications.

The second article is a work that focuses on using the power of Text Analysis and Machine learning. Adhikari et al. [192] presented 2018 a noble method for predicting disease-prone areas. The primary objective of this work is to develop an Epidemic Search model utilizing the power of social network data analysis and then employing this data to provide a probability score of the spread and to analyze the areas likely to be affected by any epidemic spread. The authors extracted tweets containing the name of the epidemic using Twitter. They have attempted to analyze and demonstrate how the model with various preprocessing techniques and algorithms predicts the output. Combining words-ngrammes, word embeddings, and TF-IDF with various data mining and deep learning algorithms such as SVM, Nave Bayes, and RNN-LSTM was utilized. As a result, Nave Bayes with TF-IDF performed superiorly compared to other algorithms.

The third review is related to the work conducted by Yeng et al. [193] in 2019. This study focused on the EDMON (Electronic Disease Monitoring Network) project, aiming to detect the early spread of contagious diseases. By employing a hybrid approach that combined K-nearness Neighbor (KNN) and cumulative sum (CUSUM) algorithms, the research achieved impressive results. The KNN algorithm demonstrated a remarkable 99.52% accuracy in classifying infected individuals,whiletheCUSUMalgorithmeffectivelyidentifiedoutbreakclusters.Thiscombinationofspatialand temporal analysis, supported by machine learning techniques and data visualization tools, showcases the potential for accurate and timely disease outbreak detection. The study's approach holds promise for proactive public health intervention and safeguarding community well-being.

Table 15.Outbreak Detection Using Hybrid Methods and Social Media or Internet Data Sources

No	Models	Data Source	Result	Limitations/ Future work
1	•Bayesian network (BN) [173] (2016) •Hidden Markov Model(HMM) •SVM •SVM-RBM	Influenza-related posts from the Sina microblog	Compared to other unsupervised and supervised models, the hybrid SVM-RBM model demonstrated greater robustness and efficacy.	The first step in future research is to determine whether the likelihood of being in an epidemic phase may be influenced by the rate from the previous week. The addition of a multivariate or spatial component to the proposed model would allow us to investigate any geographical disaggregation of the rates.
2	•SVM •Naïve Bayes •RNN-LSTM [192] (2018)	Twitter to extract the tweets with the epidemic name	NaïveBayes with TF-IDF performed better than other methods and produced a superior outcome.	Not Available
3	•KNN •CUSUM [193] (2019)	Synthetic datasets simulating health status monitoring of 297 individuals with diabetes over 12 months, including attributes like infection status, location coordinates, date/time	The K-CUSUM hybrid framework achieved 99.52% accuracy in classifying infections using KNN and accurately detected outbreak spikes through CUSUM algorithm, demonstrating	For limitation can mention challenges in achieving effective anonymization of geolocation data while preserving its utility for disease surveillance. Future work may involve exploring unsupervised

		stamps, and personal features. Additionally, the studyconsideredmobile application-generated data tocapturebloodglucose dynamics, contributing to thediseasesurveillance framework.	effective cluster detection potential for disease surveillance.	learning methods	and	assessing the system with empiricaldataforreal-world implementation.
4	•ODANN •SVM,RF,ARIMA, AutoARIMA, LSTM, Prophet [194](2021)	Twitter	ODANN demonstrates superior performance in predicting the global growth rateofCOVID-19cases compared totraditional time-series,deeplylearning and non-deep learning models.	For Future work can mention theadditionaldata sources: Incorporating other relevant data sources could further improve ODANN's performance. Alsothereal-timepredictions: Refining the model's ability to handle real-time dataiscrucialfortimely decision-making.		

The last review focuses on a work done by Chew et al. in 2021 [194], this research proposed a novel hybrid deep learning model, Optimized Data Assimilated Neural Network (ODANN), to predict the global growth rate of COVID-19 cases. By combining natural language processing (NLP) features extracted from Twitter data with data assimilation techniques, ODANN outperforms traditional time-series models, including ARIMA, RF, SVM, LSTM, AutoARIMA, and Prophet. ODANN achieves a Root Mean Squared Error (RMSE) of 0.00282 and a Mean Absolute Error (MAE) of 0.00214, demonstrating superior performance in predicting the global growth rate of COVID-19 cases. The study highlights the importance of considering social factors, such as public sentiment, in modeling infectious disease outbreaks.

5 CHALLENGES AND FUTURE DIRECTIONS

This article provides a comprehensive survey of disease outbreak detection methods that utilize diverse surveillance systemdatasources—atopicofsignificantimportance.Whilewehavepreviouslypresentedtablessummarizing limitations highlighted in prior research endeavors, here we will delve into several pivotal limitations in a broader context.

- **Background behavior and labeled data challenges:** The identification of data anomalies requires the estab-lishment of a baseline for typical behavior. What constitutes a target signal in one context, such as a seasonal influenzaepidemic,canbecomepartofthebackgroundnoiseinanother,complicatingthedualpurposeof biosurveillance systems for both detecting natural outbreaks and bioterrorism-related or pandemic diseases.

  Ensuring optimal sensitivity during and post a seasonal influenza outbreak necessitates purging bioterrorism monitoring of all seasonal influences. However, precisely timing naturally occurring disease outbreaks in specific areas remains an intricate task. The evaluation and comparison of detection algorithms suffer from a shortage of labeling. While epidemiologists can make informed estimates regarding the sequence in which people seek care, precisely measuring the time for infections to manifest specific behaviors and the subsequent interplay of these behaviors remains intricate. The highly fluid population under observation further complicates matters.

  The nonstationary, ever-evolving background behavior, influenced by population shifts, data reporting, hospital policies, and other factors, defies conventional modeling methods, perpetuating a dearth of high-quality training data.
- **Big data collection challenges and methods:** Devising algorithms capable of dissecting vast, ever-evolving, and unstructured data prevalent in social media and online content presents ongoing challenges. Algorithms designed for real-time, accurate pandemic tracking must deliver high precision with minimal time delay. Ensuring data privacy and accessibility poses additional complexities. Social media APIs grant access solely to public data, limiting the data's scope. However, making formerly private data public can attract undesirable attention and harm(such as spamming and damage to one's reputation) [195, 196]. Addressing these challenges requires expertise in epidemiology, analysis, and advanced computational skills. Conquering social media's constraints demands a meticulous methodology proficient in API system usage, managing substantial data streams, handling noisy data, and mitigating bias in data collection. Furthermore, refining disease outbreak detection accuracy hinges on the ability to pinpoint the timing and locations of message transmission. Geo-temporal disambiguation proves challenging, and text descriptions often engender significant ambiguity in place names.
- **Data accuracy and relevancy challenges:** Data and analysis quality exert a profound impact on the effectiveness of internet-based surveillance systems. Improving data quality and accuracy involves exploring methods to ascertain a user's geographical location based on profile information and language usage in texts. Addressing

sample size limitations is also critical. Internet-based surveillance systems leverage rudimentary algorithms to identify infectious disease indicators. However, as social network usage surges and data grows exponentially, sophisticated algorithms are imperative for real-time pandemic extraction, analysis, detection, and tracking with unparalleled accuracy. Analyzing live, massive, and unstructured data mandates advanced computational linguistics. The systematic, generic approach of internet-driven surveillance renders it adaptable for monitoring a spectrum of infectious diseases. There's substantial headroom for progress in areas like detection (via high-filtermethods),testing(asexemplifiedbyZika),andintegrationwithtraditionalsurveillancesystems.To comprehensively fathom pandemic dynamics, flexible internet-based surveillance frameworks must unify data from diverse online platforms[197, 198].

- **Internet Search Context Interpretation:** Interpreting the search context of specific queries or documents in internet-based data systems is challenging due to varied user motives—queries may pertain to drugs, symptoms, or illnesses for diverse reasons. The multifaceted meanings of a single word depending on its contextual use in surrounding text introduce complexity. Furthermore, a single disease may be referred to by multiple names and exhibit a wide range of symptoms. Future research should strive to minimize false alarms and the detection of insignificant events by epidemic systems.
- **Incorporating Spatial Information and Cluster Detection using Spatiotemporal Analysis:** Amidst the myriadchallengesofdiseaseoutbreakdetection,anemergingavenueofexplorationliesintheintegration ofspatialinformationandclusterdetectionthroughspatiotemporalanalysis.Harnessingthegeographical context of outbreaks can offer invaluable insights into disease spread patterns, aiding in the identification of localized clusters and hotspots. Such spatially aware methodologies provide a holistic view of epidemic dynamics, accounting for the interplay between geographical proximity, population density, and disease transmission.

  Cluster detection techniques, particularly those rooted in spatiotemporal analysis, hold promise in unveiling hidden patterns within outbreak data. By identifying clusters of cases occurring within specific regions and timeframes, these methods offer a proactive approach to pinpointing disease emergence and spread. Spatiotem-poral cluster detection not only aids in early warning systems but also facilitates targeted intervention strategies, optimizing resource allocation and containment efforts. As the field advances, combining spatial information with machine learning algorithms becomes increasingly

  pertinent.Leveragingadvancedcomputationaltechniques,suchasgeospatialmachinelearning,holdsthe potential to unravel complex relationships between disease dynamics, environmental factors, and human behavior.

  Bydevelopingmodelsthatfusetemporaltrends,geographicalproximity,andsociodemographicvariables, researchers can attain a nuanced understanding of outbreak propagation.
- **Challenges and Future Directions in Spatial-Based Detection:** However, incorporating spatial information introduces its own set of challenges. Spatial data often exhibits heterogeneity and uneven distribution, necessitat-ing careful preprocessing and normalization. Additionally, reconciling disparate data sources, such as healthcare facility records, geolocation data, and social media feeds, requires sophisticated data fusion techniques. Effective utilization of spatial information also calls for domain expertise in geospatial analysis and disease epidemiology.

  Future research endeavors can focus on refining the accuracy and reliability of spatiotemporal cluster detection methods. This involves developing algorithms that account for varying spatial scales, temporal resolutions, and the incorporation of uncertainty measures. By integrating real-time data streams from remote sensing platforms, mobile devices, and wearable technology, researchers can enhance the timeliness and granularity of outbreak alerts.

Furthermore, ethical considerations must accompany the utilization of spatial data, especially when dealing with sensitive information about individuals' locations and movements. Striking a balance between public health interests and individual privacy remains a critical aspect of spatial-based surveillance.

- **Emerging Challenges and Areas of Exploration:** While our survey comprehensively covers existing chal-lenges, it's important to acknowledge that the field is ever-evolving. Emerging challenges may include issues related to cross-platform data integration, the ethical use of user-generated content, and the continuous re-finement of algorithmic accuracy. Moreover, considering the rapid pace of technological advancements, new opportunitiesforharnessingnoveldatasourcesandemployingadvancedanalyticaltechniquesmayarise, reshaping the landscape of disease outbreak detection.
  By illuminating these intricate challenges, this survey equips researchers, practitioners, and policymakers with valuable insights as they navigate the dynamic realm of disease outbreak detection. As this field continues to evolve, this survey's relevance and impact are poised to endure.
- **Embracing Diverse Data Sources:** In light of the evolving landscape of data-driven research for outbreak detection,itisworthconsideringanexpandedexplorationofdatasourcesbeyondthedichotomyofsocial media/internet versus conventional data sources. As the field progresses, integrating emerging data streams such as satellite imagery, healthcare claims, retail purchases, environmental sensors, and genomic sequencing could offer a more comprehensive and dynamic approach to early outbreak detection. Additionally, datasets from platforms like YouTube, LinkedIn, and facebook, should be explored, especially considering the increasing limitationsonTwitterAPIaccessibility.Thesealternativesocialmediadatasourcescouldprovidevaluable insights, further enhancing the diversity of inputs available for predictive modeling. This inclusive perspective acknowledges the multifaceted nature of outbreaks and leverages a diverse array of data sources to enhance accuracy, timeliness, and reliability in predicting and responding to health threats. Furthermore, ensuring the privacy and ethical considerations surrounding the collection and utilization of all these data sources remains a critical challenge that requires careful attention and robust solutions.

6 CONCLUSION

This article presents a comprehensive overview of conventional and internet-based surveillance systems for identifying disease outbreaks and forecasting, offering an invaluable resource for researchers and practitioners alike. Our analysis coversawidespectrumofliterature,focusingonthekeydimensionsofdatasources,methodologies,algorithmic comparisons, challenges, and future directions. By scrutinizing historical trends in the field, we aim to capture the evolution of outbreak detection techniques and forecasting, with a particular focus on data-driven approaches. Our review is structured around two pivotal axes: (1) data sources, and (2) statistical and machine learning techniques. A key finding of this survey is that 62% of the studies utilized conventional data sources, while 38% relied on social media or internet-derived data. Additionally, 20% of the research employed statistical models, 28% leveraged machine learning models (including both deep learning and non-deep learning techniques), and 16% applied hybrid models. These insights underline the growing adoption of machine learning methods, alongside the continued relevance of traditional statistical approaches, in outbreak prediction and forecasting. This survey not only highlights the ongoing exploration of diverse frameworks and methodologies but also brings attention to the underutilized potential of hybrid and emerging models. Several critical factors influencing the performance of outbreak detection were identified, including data origin, dataset size, real-time data accessibility, machine learning hyperparameters, geolocation effects, and healthcare facility variables. Addressing these determinants, alongside the limitations and challenges outlined in the existing literature, will be

crucial for future research. Our survey aims to close existing gaps by providing a cohesive synthesis of these methods

and by suggesting potential directions for innovation.

The findings and discussions in this article will aid academics, researchers, clinicians, healthcare practitioners, and policymakers in navigating the evolving landscape of public health applications. As the domain of disease outbreak detection continues to expand, we hope this up-to-date survey fosters the exchange of ideas, encourages the development

of novel techniques, and serves as a vital resource for shaping future research.

REFERENCES

[1]V.Aakash,S.Sridevi,G.Ananthi,andS.Rajaram. ForecastingofNovelCoronaVirusDisease(Covid-19)UsingLSTMandXG BoostingAlgorithms. In*DataAnalyticsinBioinformatics*,pages293–311.JohnWiley&Sons,Ltd,2021. Section:12_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119785620.ch12.

[2]Emily H. Chan, Vikram Sahai, Corrie Conrad, and John S. Brownstein.Using Web Search Query Data to Monitor Dengue Epidemics: A New Model for Neglected Tropical Disease Surveillance.*PLOS Neglected Tropical Diseases*, 5(5):e1206, May 2011.Publisher: Public Library of Science. [3]Y. Tony Yang, Michael Horneffer, and Nicole DiLisio.Mining Social Media and Web Searches for Disease Detection.*Journal of Public Health Research*, 2(1):jphr.2013.e4, March 2013.

[4]Cédric Abat, Hervé Chaudet, Jean-Marc Rolain, Philippe Colson, and Didier Raoult.Traditional and syndromic surveillance of infectious diseases and pathogens.*International Journal of Infectious Diseases*, 48:22–28, 2016.

[5]Abid Haleem, Mohd Javaid, Ravi Pratap Singh, and Rajiv Suman.Telemedicine for healthcare: Capabilities, features, barriers, and applications. *Sensors international*, 2:100117, 2021.

[6]Elham Monaghesh and Alireza Hajizadeh.The role of telehealth during covid-19 outbreak: a systematic review based on current evidence.*BMC public health*, 20:1–9, 2020.

[7]Eirini Christaki.New technologies in predicting, preventing and controlling emerging infectious diseases.*Virulence*, 6(6):558–565, August 2015. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21505594.2015.1040975.

[8]Mauricio Santillana, André T. Nguyen, Mark Dredze, Michael J. Paul, Elaine O. Nsoesie, and John S. Brownstein.Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance.*PLOS Computational Biology*, 11(10):e1004513, October 2015.Publisher: Public Library of Science.

[9]Ed De Quincey and Patty Kostkova. Early warning and outbreak detection using social networking websites: The potential of twitter. In *International conference on electronic healthcare*, pages 21–24. Springer, 2009.

[10]Khaled Al-Surimi, Mohammed Khalifa, Salwa Bahkali, Ashraf El-Metwally, and Mowafa Househ.The potential of social media and internet-based data in preventing and fighting infectious diseases: from internet to twitter.In *Emerging and Re-emerging Viral Infections*, pages 131–139. Springer, 2016.

[11]Yusheng Xie, Zhengzhang Chen, Cheng, Yu, Zhang, Kunpeng, Agrawal, Ankit, Liao, Wei-keng, and Choudhary, Alok.Detecting and Tracking Disease Outbreaks by Mining Social Media Data.*Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, page 3, 2013.

[12]Erik Bohlin.*Tracking the outbreak of diseases Using Twitter : A Machine Learning Approach*.2012.

[13]Vasileios Lampos, Andrew C Miller, Steve Crossan, and Christian Stefansen.Advances in nowcasting influenza-like illness rates using search query logs.*Scientific reports*, 5(1):1–10, 2015.

[14]Jens P Linge, Ralf Steinberger, TP Weber, Roman Yangarber, Erik van der Goot, DH Al Khudhairy, and NI Stilianakis.Internet surveillance systems for early alerting of health threats.*Eurosurveillance*, 14(13):19162, 2009.

[15]Ed De Quincey and Patty Kostkova.Early warning and outbreak detection using social networking websites: The potential of twitter.In *Electronic Healthcare: Second International ICST Conference, eHealth 2009, Istanbul, Turkey, September 23-15, 2009, Revised Selected Papers 2*, pages 21–24. Springer, 2010.

[16]Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant.Detecting influenza epidemics using search engine query data.*Nature*, 457(7232):1012–1014, 2009.

[17]Cynthia Chew and Gunther Eysenbach.Pandemics in the age of twitter: content analysis of tweets during the 2009 h1n1 outbreak.*PloS one*, 5(11):e14118, 2010.

[18]Lawrence C Madoff and John P Woodall. The internet and the global monitoring of emerging diseases: lessons from the first 10 years of promed-mail. *Archives of medical research*, 36(6):724–730, 2005.

[19]Gema Bello-Orgaz, Julio Hernandez-Castro, and David Camacho.A Survey of Social Web Mining Applications for Disease Outbreak Detection.In David Camacho, Lars Braubach, Salvatore Venticinque, and Costin Badica, editors, *Intelligent Distributed Computing VIII*, Studies in Computational Intelligence, pages 345–356, Cham, 2015. Springer International Publishing.

[20]Mohammed Ali Al-garadi, Muhammad Sadiq Khan, Kasturi Dewi Varathan, Ghulam Mujtaba, and Abdelkodose M. Al-Kabsi.Using online social networks to track a pandemic: A systematic review.*Journal of Biomedical Informatics*, 62:1–11, August 2016.

[21]Harald Hornmoen and Colin McInnes.Social Media Communication During Disease Outbreaks: Findings and Recommendations.In Harald Hornmoen and Klas Backholm, editors, *Social Media Use in Crisis and Risk Communication*, pages 255–275. Emerald Publishing Limited, October 2018.
[22]Eunjoo Yang, Hyun Park, Yeon Choi, Jusim Kim, Lkhagvadorj Munkhdalai, Ibrahim Musa, and Keun Ryu.A Simulation-Based Study on the Comparison of Statistical and Time Series Forecasting Methods for Early Detection of Infectious Disease Outbreaks.*IJERPH*, 15(5):966, May 2018. [23]ChrisChatfieldandMohammadYar.Holt-WintersForecasting:SomePracticalIssues.*JournaloftheRoyalStatisticalSociety:SeriesD(The Statistician)*, 37(2):129–140, 1988._eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2307/2348687.
[24]GuohunZhu,LipingLi,YuebinZheng,XiaoweiZhang,andHuiZou.ForecastingInfluenzaBasedonAutoregressiveMovingAverageand Holt-Winters Exponential Smoothing Models.*Journal of Advanced Computational Intelligence and Intelligent Informatics*, 25(1):138–144, 2021. [25]Mrutyunjaya Panda.Application of ARIMA and Holt-Winters forecasting model to predict the spreading of COVID-19 for India and its states, July 2020.Pages: 2020.07.14.20153908.
[26]Seng Hansun, Vincent Charles, Tatiana Gherman, Subanar, and Christiana Rini Indrati.A tuned Holt-Winters white-box model for COVID-19 prediction.*International Journal of Management and Decision Making*, 20(3):241–262, January 2021.Publisher: Inderscience Publishers.
[27]Norwaziah Mahmud, Nur Syuhada Muhammat Pazil, Hafawati Jamaluddin, and Nur Aqilah Ali.Prediction of dengue outbreak: a comparison between ARIMA and holt-winters methods / Norwaziah Mahmud ... [et al.].*ESTEEM Academic Journal*, 17:101–111, August 2021.Publisher: Universiti Teknologi MARA Cawangan Pulau Pinang.
[28]Yuzhou Zhang, Gabriel Milinovich, Zhiwei Xu, Hilary Bambrick, Kerrie Mengersen, Shilu Tong, and Wenbiao Hu.Monitoring pertussis infections using internet search queries.*Scientific reports*, 7(1):1–7, 2017.
[29]I Djakaria and SE Saleh.Covid-19 forecast using holt-winters exponential smoothing.In *Journal of Physics: Conference Series*, volume 1882, page 012033. IOP Publishing, 2021.
[30]Sarab D. Shukur and Tasnim Hasan Kadhim.Time series analysis of the number of Covid-19 deaths in Iraq.*International Journal of Nonlinear Analysis and Applications*, 12(2):1997–2007, July 2021.Publisher: Semnan University.
[31]S. S. Wickramasinghe and K. M. U. B. Konarasinghe.Forecasting COVID-19 Daily Infected Cases in Sri Lanka by Holt-Winters Model.March 2022. Accepted: 2022-04-22T05:43:46Z Publisher: University of Ruhuna, Matara, Sri Lanka.
[32]Simon James Fong, Gloria Li, Nilanjan Dey, Rubén González Crespo, and Enrique Herrera-Viedma.Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak.*International Journal of Interactive Multimedia and Artificial Intelligence*, 6, March 2020.
[33]Arul Earnest, Mark I. Chen, Donald Ng, and Leo Yee Sin.Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore.*BMC Health Services Research*, 5(1):36, May 2005. [34]Debabrata Dansana, Raghvendra Kumar, Janmejoy Das Adhikari, Mans Mohapatra, Rohit Sharma, Ishaani Priyadarshini, and Dac-Nhuong Le. Global Forecasting Confirmed and Fatal Cases of COVID-19 Outbreak Using Autoregressive Integrated Moving Average Model.*Frontiers in Public Health*, 8, 2020.
[35]HamidRezaPourghasemi,SoheilaPouyan,ZakariyaFarajzadeh,NitheshnirmalSadhasivam,BahramHeidari,SedighehBabaei,andJohnP. Tiefenbacher.Assessment of the outbreak risk, mapping and infection behavior of COVID-19: Application of the autoregressive integrated-moving average (ARIMA) and polynomial models.*PLOS ONE*, 15(7):e0236238, July 2020.Publisher: Public Library of Science.
[36]Stefan H. Steiner, Kristina Grant, Michael Coory, and Heath A. Kelly.Detecting the start of an influenza outbreak using exponentially weighted moving average charts.*BMC Medical Informatics and Decision Making*, 10(1):37, June 2010.
[37]Sameh Nassar, Klaus-Peter Schwarz, Naser El-Sheimy, and Aboelmagd Noureldin.Modeling Inertial Sensor Errors Using Autoregressive (AR) Models.*NAVIGATION*, 51(4):259–268, 2004._eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2161-4296.2004.tb00357.x.
[38]Aman Khakharia, Vruddhi Shah, Sankalp Jain, Jash Shah, Amanshu Tiwari, Prathamesh Daphal, Mahesh Warang, and Ninad Mehendale.Outbreak Prediction of COVID-19 for Dense and Populated Countries Using Machine Learning.*Ann. Data. Sci.*, 8(1):1–19, March 2021.
[39]Yuzhou Zhang, Hilary Bambrick, Kerrie Mengersen, Shilu Tong, and Wenbiao Hu.Using Google Trends and ambient temperature to predict seasonal influenza outbreaks.*Environment International*, 117:284–291, August 2018.
[40]Nalini Chintalapudi, Gopi Battineni, and Francesco Amenta.COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach.*Journal of Microbiology, Immunology and Infection*, 53(3):396–403, June 2020.
[41]Paloma Monllor, Zhenyu Su, Laura Gabrielli, and Paloma Taltavull de La Paz.COVID-19 Infection Process in Italy and Spain: Are Data Talking? Evidence From ARMA and Vector Autoregression Models.*Frontiers in Public Health*, 8, 2020.
[42]KhairanRajab,FiruzKamalov,andAswaniKumarCherukuri.ForecastingCOVID-19:VectorAutoregression-BasedModel.*ArabJSciEng*, 47(6):6851–6860, June 2022.
[43]Antonio Oliva, Francesco Gracceva, Daniele Lerede, Matteo Nicoli, and Laura Savoldi.Projection of Post-Pandemic Italian Industrial Production through Vector AutoRegressive Models.*Energies*, 14(17):5458, January 2021.Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.
[44]Qinan Wang, Yaomu Zhou, and Xiaofei Chen.A Vector Autoregression Prediction Model for COVID-19 Outbreak.February 2021.
[45]Helmut Lütkepohl.Vector autoregressive models.*Handbook of Research Methods and Applications in Empirical Macroeconomics*, pages 139–164, July 2013.ISBN: 9780857931023 Publisher: Edward Elgar Publishing Section: Handbook of Research Methods and Applications in Empirical Macroeconomics.

[46]Rochelle E. Watkins, Serryn Eagleson, Bert Veenendaal, Graeme Wright, and Aileen J. Plant.Applying cusum-based methods for the detection of outbreaks of Ross River virus disease in Western Australia.*BMC Med Inform Decis Mak*, 8(1):37, August 2008.
[47]K. Sharifolkashani, P. Yavari, R. Shekarriz, F. Tajdini, and N. Aghili.Early Detection of Dysentery Outbreaks by Cumulative Sum Method Based on National Surveillance System Data in 1393-1396.*Iranian Journal of Epidemiology*, 16(4):276–284, March 2021.Publisher: Iranian Journal of Epidemiology.
[48]Manoochehr KARAMI, Maryam GHALANDARI, Jalal POOROLAJAL, and Javad FARADMAL. Early Detection of Meningitis Outbreaks: Application of Limited-baseline Data.*Iran J Public Health*, 46(10):1366–1373, October 2017.
[49]Richard John M. Buendia and Geoffrey A. Solano.A disease outbreak detection system using autoregressive moving average in time series analysis. In *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pages 1–5, July 2015.
[50]Christopher J. Lynch and Ross Gore.Application of one-, three-, and seven-day forecasts during early onset on the COVID-19 epidemic dataset using moving average, autoregressive, autoregressive moving average, autoregressive integrated moving average, and naïve forecasting methods. *Data in Brief*, 35:106759, April 2021.
[51]Ram Kumar Singh, Meenu Rani, Akshaya Srikanth Bhagavathula, Ranjit Sah, Alfonso J. Rodriguez-Morales, Himangshu Kalita, Chintan Nanda, Shashi Sharma, Yagya Datt Sharma, Ali A. Rabaan, Jamal Rahmani, and Pavan Kumar.Prediction of the COVID-19 Pandemic for the Top 15 Affected Countries: Advanced Autoregressive Integrated Moving Average (ARIMA) Model.*JMIR Public Health and Surveillance*, 6(2):e19115, May 2020.Company: JMIR Public Health and Surveillance Distributor: JMIR Public Health and Surveillance Institution: JMIR Public Health and Surveillance Label: JMIR Public Health and Surveillance Publisher: JMIR Publications Inc., Toronto, Canada.
[52]Gülhan Toğa, Berrin Atalay, and M. Duran Toksari.COVID-19 prevalence forecasting using Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN): Case of Turkey.*Journal of Infection and Public Health*, 14(7):811–816, July 2021.
[53]K. E. ArunKumar, Dinesh V. Kalaga, Ch. Mohan Sai Kumar, Govinda Chilkoor, Masahiro Kawaji, and Timothy M. Brenza.Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA).*Applied Soft Computing*, 103:107161, May 2021.
[54]Vahid Rahmanian, Saied Bokaie, Aliakbar Haghdoost, and Mohsen Barouni.Predicting cutaneous leishmaniasis using sarima and markov switching models in isfahan, iran: A time-series study.*Asian Pacific Journal of Tropical Medicine*, 14(2):83–93, 2021.
[55]Rubén Amorós, David Conesa, Antonio López-Quílez, and Miguel-Angel Martinez-Beneito.A spatio-temporal hierarchical Markov switching model for the early detection of influenza outbreaks.*Stoch Environ Res Risk Assess*, 34(2):275–292, February 2020.
[56]RubénAmorósSalvador. *BayesiantemporalandspatioͲtemporalMarkovswitchingmodelsforthedetectionofinfluenzaoutbreaks*. http://purl.org/dc/dcmitype/Text, Universitat de València, 2017.
[57]Hsin-Min Lu, Daniel Zeng, and Hsinchun Chen.Bioterrorism event detection based on the Markov switching model: A simulated anthrax outbreak study.In *2008 IEEE International Conference on Intelligence and Security Informatics*, pages 76–81, June 2008.
[58]Francesco Bartolucci and Alessio Farcomeni.A spatio-temporal model based on discrete latent variables for the analysis of covid-19 incidence. *Spatial Statistics*, 49:100504, 2022.
[59]Navid Feroze, Kamran Abbas, Farzana Noor, and Amjad Ali.Analysis and forecasts for trends of covid-19 in pakistan using bayesian models.*PeerJ*, 9:e11537, 2021.
[60]DCostagliola,AFlahault,DGalinec,PGarnerin,JMenares,andAJValleron.Aroutinetoolfordetectionandassessmentofepidemicsof influenza-like syndromes in France.*Am J Public Health*, 81(1):97–99, January 1991.Publisher: American Public Health Association.
[61]R. Snacken, J. Lion, V. Van Casteren, R. Cornelis, F. Yane, M. Mombaerts, W. Aelvoet, and A. Stroobant.Five years of sentinel surveillance of acute respiratory infections (1985–1990): The benefits of an influenza early warning system.*Eur J Epidemiol*, 8(4):485–490, July 1992.
[62]Donna F. Stroup and Stephen B. Thacker.A Bayesian Approach to the Detection of Aberrations in Public Health Surveillance Data.*Epidemiology*, 4(5):435–443, 1993.Publisher: Lippincott Williams & Wilkins.
[63]KEEWHAN CHOI and STEPHEN B. THACKER.AN EVALUATION OF INFLUENZA MORTALITY SURVEILLANCE, 1962–1979: I. TIME SERIES FORECASTS OF EXPECTED PNEUMONIA AND INFLUENZA DEATHS.*American Journal of Epidemiology*, 113(3):215–226, March 1981. [64]Donna F. Stroup, Melinda Wharton, Karen Kafadar, and Andrew G. Dean.Evaluation of a Method for Detecting Aberrations in Public Health Surveillance Data.*American Journal of Epidemiology*, 137(3):373–380, February 1993.
[65]Andrew D. Cliff, Peter Haggett, Donna F. Stroup, and Elizabeth Cheney.The changing geographical coherence of measles morbidity in the United States, 1962–88.*Statistics in Medicine*, 11(11):1409–1424, 1992._eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.4780111102.
[66]FLAVIO F NOBRE and DONNA F STROUP.A Monitoring System to Detect Changes in Public Health Surveillance Data.*International Journal of Epidemiology*, 23(2):408–418, April 1994.
[67]Santanu Roy, Gouri Sankar Bhunia, and Pravat Kumar Shit.Spatial prediction of covid-19 epidemic using arima techniques in india.*Modeling earth systems and environment*, 7:1385–1391, 2021.
[68]Samir I. Thaker, Amy S. Nowacki, Neil B. Mehta, and Ashley R. Edwards.How U.S. Hospitals Use Social Media.*Ann Intern Med*, 154(10):707–708, May 2011.Publisher: American College of Physicians.
[69]Ying Chen, Yuzhou Zhang, Zhiwei Xu, Xuanzhuo Wang, Jiahai Lu, and Wenbiao Hu.Avian influenza a (h7n9) and related internet search query data in china.*Scientific reports*, 9(1):10434, 2019.

[70]Loukas Samaras, Elena García-Barriocanal, and Miguel-Angel Sicilia.Comparing Social media and Google to detect and predict severe epidemics. *Sci Rep*, 10(1):4747, December 2020.
[71]Lei Qin, Qiang Sun, Yidan Wang, Ke-Fei Wu, Mingchih Chen, Ben-Chang Shia, and Szu-Yuan Wu.Prediction of Number of Cases of 2019 Novel Coronavirus (COVID-19) Using Social Media Search Index.*International Journal of Environmental Research and Public Health*, 17(7):2365, January 2020.Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.
[72]Muhammed Navas Thorakkattle, Shazia Farhin, and Athar Ali Khan.Forecasting the trends of covid-19 and causal impact of vaccines using bayesian structural time series and arima.*Annals of Data Science*, 9(5):1025–1047, 2022.
[73]Amy Ming-Fang Yen, Tony Hsiu-Hsi Chen, Wei-Jung Chang, Ting-Yu Lin, Grace Hsiao-Hsuan Jen, Chen-Yang Hsu, Sen-Te Wang, Huong Dang, and Sam Li-Sheng Chen.New surveillance metrics for alerting community-acquired outbreaks of emerging sars-cov-2 variants using imported case data: Bayesian markov chain monte carlo approach.*JMIR Public Health and Surveillance*, 8(11):e40866, 2022.
[74]Amir Hassan Zadeh, Hamed M Zolbanin, Ramesh Sharda, and Dursun Delen.Social media for nowcasting flu activity: Spatio-temporal big data analysis.*Information Systems Frontiers*, 21:743–760, 2019.
[75]Rubén Amorós, David Conesa, Antonio López-Quílez, and Miguel-Angel Martinez-Beneito.A spatio-temporal hierarchical markov switching model for the early detection of influenza outbreaks.*Stochastic Environmental Research and Risk Assessment*, 34(2):275–292, 2020.
[76]Zhensheng Wang, Yang Yue, Biao He, Ke Nie, Wei Tu, Qingyun Du, and Qingquan Li.A bayesian spatio-temporal model to analyzing the stability of patterns of population distribution in an urban space using mobile phone data.*International Journal of Geographical Information Science*, 35(1):116–134, 2021.
[77]W Suryaningrat, D Munandar, A Maryati, AS Abdullah, and BN Ruchjana.Posted prediction in social media base on markov chain model: twitter dataset with covid-19 trends.In *Journal of Physics: Conference Series*, volume 1722, page 012001. IOP Publishing, 2021.
[78]SPradeepaandKRManjula.Epidemiczoneofcovid-19fromsocialmediausinghypergraphwithweightingfactor(hwf).*TheJournalof Supercomputing*, 77:11738–11755, 2021.
[79]YuanShi,XuLiu,Suet-YhengKok,JayanthiRajarethinam,ShaohongLiang,GraceYap,Chee-SengChong,Kim-SungLee,SharonS.Y.Tan, Christopher Kuan Yew Chin, Andrew Lo, Waiming Kong, Lee Ching Ng, and Alex R. Cook.Three-Month Real-Time Dengue Forecast Models: An Early Warning System for Outbreak Alerts and Policy Decision Support in Singapore.*Environmental Health Perspectives*, 124(9):1369–1375, September 2016.Publisher: Environmental Health Perspectives.
[80]Sui Lan Tang and Preethi Subramanian.Review on Nowcasting using Least Absolute Shrinkage Selector Operator (LASSO) to Predict Dengue Occurrence in San Juan and Iquitos as Part of Disease Surveillance System.*Periodicals of Engineering and Natural Sciences*, 7(2):608–617, July 2019. Number: 2.
[81]Furqan Rustam, Aijaz Ahmad Reshi, Arif Mehmood, Saleem Ullah, Byung-Won On, Waqar Aslam, and Gyu Sang Choi. COVID-19 Future Forecasting Using Supervised Machine Learning Models.*IEEE Access*, 8:101489–101499, 2020.Conference Name: IEEE Access.
[82]Pi Guo, Tao Liu, Qin Zhang, Li Wang, Jianpeng Xiao, Qingying Zhang, Ganfeng Luo, Zhihao Li, Jianfeng He, Yonghui Zhang, and Wenjun Ma. Developing a dengue forecast model using machine learning: A case study in China.*PLOS Neglected Tropical Diseases*, 11(10):e0005973, October 2017.Publisher: Public Library of Science.
[83]Tsair-Fwu Lee, Pei-Ju Chao, Hui-Min Ting, Liyun Chang, Yu-Jie Huang, Jia-Ming Wu, Hung-Yu Wang, Mong-Fong Horng, Chun-Ming Chang, Jen-Hong Lan, Ya-Yu Huang, Fu-Min Fang, and Stephen Wan Leung.Using Multivariate Regression Model with Least Absolute Shrinkage and Selection Operator (LASSO) to Predict the Incidence of Xerostomia after Intensity-Modulated Radiotherapy for Head and Neck Cancer.*PLOS ONE*, 9(2):e89700, February 2014.Publisher: Public Library of Science.
[84]Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani.Least angle regression.*The Annals of Statistics*, 32(2):407–499, April 2004. Publisher: Institute of Mathematical Statistics.
[85]Qing Wang, Mo Bai, and Mai Huang.Empirical Examination on the Drivers of the U.S. Equity Returns in the During the COVID-19 Crisis.*Frontiers in Public Health*, 9, 2021.
[86]Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, et al.Xgboost: extreme gradient boosting.*R package version 0.4-2*, 1(4):1–4, 2015.
[87]Tianqi Chen and Carlos Guestrin.XGBoost: A Scalable Tree Boosting System.In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, August 2016. Association for Computing Machinery. [88]Jouhyun Jeon, Peter J. Leimbigler, Gaurav Baruah, Michael H. Li, Yan Fossat, and Alfred J. Whitehead.Predicting Glycaemia in Type 1 Diabetes Patients: Experiments in Feature Engineering and Data Imputation.*J Healthc Inform Res*, 4(1):71–90, March 2020.
[89]Kumar Shashvat, Rikmantra Basu, Amol P. Bhondekar, and Arshpreet Kaur.Epidemiology and Forecasting of Cholera Incidence in North India.In Vinit Kumar Gunjan, Sabrina Senatore, Amit Kumar, Xiao-Zhi Gao, and Suresh Merugu, editors, *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering, pages 9–17. Springer, Singapore, 2020.
[90]Mahrukh Saif, Muhammad Asif Zahoor Raja, and Aneela Zameer.Analysis of Covid-19 Literature Evolution via NLP and Machine Learning.In *2022 International Conference on Recent Advances in Electrical Engineering & Computer Sciences (RAEE & CS)*, pages 1–8, October 2022.
[91]Dr M Lalli.Optimized Deep Learning Based Ensemble Model for Forecasting of Covid-19.13:6, 2021.
[92]Godson Kalipe, Vikas Gautham, and Rajat Kumar Behera.Predicting Malarial Outbreak using Machine Learning and Deep Learning Approach: A Review and Analysis.In *2018 International Conference on Information Technology (ICIT)*, pages 33–38, December 2018.

[93]Zheng-gang Fang, Shu-qin Yang, Cai-xia Lv, Shu-yi An, and Wei Wu.Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study.*BMJ Open*, 12(7):e056685, July 2022.Publisher: British Medical Journal Publishing Group Section: Epidemiology. [94]Rohil Badkundri, Victor Valbuena, Srikusmanjali Pinnamareddy, Brittney Cantrell, and Janet Standeven.Forecasting the 2017-2018 Yemen Cholera Outbreak with Machine Learning, February 2019.arXiv:1902.06739 [cs, q-bio].
[95]Vijander Singh, Ramesh Chandra Poonia, Sandeep Kumar, Pranav Dass, Pankaj Agarwal, Vaibhav Bhatnagar, and Linesh Raja.Prediction of COVID-19 corona virus pandemic based on time series data using support vector machine.*Journal of Discrete Mathematical Sciences and Cryptography*, 23(8):1583–1597, November 2020.Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09720529.2020.1784535.
[96]Amit Kumar Gupta, Vijander Singh, Priya Mathur, and Carlos M. Travieso-Gonzalez.Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario.*Journal of Interdisciplinary Mathematics*, 24(1):89–108, January 2021.Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09720502.2020.1833458.
[97]Vikramaditya Jakkula.Tutorial on support vector machine (svm).*School of EECS, Washington State University*, 37(2.5):3, 2006.
[98]Gurcan Comert, Negash Begashaw, and Ayse Turhan-Comert.Malaria outbreak detection with machine learning methods.*BioRxiv*, 2020. [99]NorFarishaMuhamadKrishnan,ZurianiAhmadZukarnain,AzlinAhmad,andMarhainisJamaludin.PredictingDengueOutbreakbased on Meteorological Data Using Artificial Neural Network and Decision Tree Models.*JOIV : International Journal on Informatics Visualization*, 6(3):597–603, September 2022.Number: 3.
[100]Md. Ashikur Rahman Khan, Jony Akter, Ishtiaq Ahammad, Sabbir Ejaz, and Tanvir Jaman Khan.Dengue outbreaks prediction in Bangladesh perspective using distinct multilayer perceptron NN and decision tree.*Health Inf Sci Syst*, 10(1):32, November 2022.
[101]Hakizimana Leopord, W Kipruto Cheruiyot, and Stephen Kimani.A survey and analysis on classification and regression data mining techniques for diseases outbreak prediction in datasets.*Int. J. Eng. Sci*, 5(9):1–11, 2016.
[102]T. Lowie, J. Callens, J. Maris, S. Ribbens, and B. Pardon.Decision tree analysis for pathogen identification based on circumstantial factors in outbreaks of bovine respiratory disease in calves.*Preventive Veterinary Medicine*, 196:105469, November 2021.
[103]Vili Podgorelec, Peter Kokol, Bruno Stiglic, and Ivan Rozman.Decision Trees: An Overview and Their Use in Medicine.*Journal of Medical Systems*, 26(5):445–463, October 2002.
[104]Liaqat Ali, Shafqat Ullah Khan, Noorbakhsh Amiri Golilarz, Imrana Yakubu, Iqbal Qasim, Adeeb Noor, and Redhwan Nour.A Feature-Driven Decision Support System for Heart Failure Prediction Based on Statistical Model and Gaussian Naive Bayes.*Computational and Mathematical Methods in Medicine*, 2019:e6314328, November 2019.Publisher: Hindawi.
[105]H. Zakiyyah and S. Suyanto.Prediction of Covid-19 Infection in Indonesia Using Machine Learning Methods.*J. Phys.: Conf. Ser.*, 1844(1):012002, March 2021.Publisher: IOP Publishing.
[106]Eric Yunan Zhao, Daniel Xia, Mark Greenhalgh, Elena Colicino, Merylin Monaro, Rita Hitching, Odette A. Harris, and Maheen M. Adamson. Combining International Survey Datasets to Identify Indicators of Stress during the COVID-19 Pandemic: A Machine Learning Approach to Improve Generalization.*COVID*, 1(4):728–738, December 2021.Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
[107]Seyed Masoud Rezaeijo, Razzagh Abedi-Firouzjah, Mohammadreza Ghorvei, and Samad Sarnameh.Screening of COVID-19 based on the extracted radiomics features from chest CT images.*Journal of X-Ray Science and Technology*, 29(2):229–243, January 2021.Publisher: IOS Press.
[108]Zanya Reubenne D. Omadlao, Johanna Marie A. Cabrales, Samuel Christian M. Cristobal, Margaret Vianey A. Dee, Jim Reinier V. Tadeo, Joseph Ludwin D. C. Marigmen, and Romsto R. Pajarillo.Machine learning-based dengue forecasting system for Irisan, Baguio city, Philippines.*AIP Conference Proceedings*, 2472(1):040019, August 2022.Publisher: American Institute of Physics.
[109]Michael J. Kane, Natalie Price, Matthew Scotch, and Peter Rabinowitz.Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks.*BMC Bioinformatics*, 15(1):276, August 2014.
[110]Ruirui Liang, Yi Lu, Xiaosheng Qu, Qiang Su, Chunxia Li, Sijing Xia, Yongxin Liu, Qiang Zhang, Xin Cao, Qin Chen, and Bing Niu.Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data.*Transboundary and Emerging Diseases*, 67(2):935–946, 2020._eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tbed.13424.
[111]Debabrata Dansana, Raghvendra Kumar, Aishik Bhattacharjee, and Chandrakanta Mahanty.COVID-19 Outbreak Prediction and Analysis of E-Healthcare Data Using Random Forest Algorithms.*IJRQEH*, 11(1):1–13, January 2022.Publisher: IGI Global.
[112]Cafer Mert Yeşilkanat.Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm. *Chaos, Solitons & Fractals*, 140:110210, November 2020.
[113]Janet Ong, Xu Liu, Jayanthi Rajarethinam, Suet Yheng Kok, Shaohong Liang, Choon Siang Tang, Alex R. Cook, Lee Ching Ng, and Grace Yap. Mapping dengue risk in Singapore using Random Forest.*PLOS Neglected Tropical Diseases*, 12(6):e0006587, June 2018.Publisher: Public Library of Science.
[114]Yirong Chen, Collins Wenhan Chu, Mark I. C. Chen, and Alex R. Cook.The utility of LASSO-based models for real time forecasts of endemic infectious diseases: A cross country comparison.*Journal of Biomedical Informatics*, 81:16–30, May 2018.
[115]Shanen Chen, Jian Xu, Yongsheng Wu, Xin Wang, Shisong Fang, Jinquan Cheng, Hanwu Ma, Renli Zhang, Yachuan Liu, Li Zhang, et al.Predicting temporal propagation of seasonal influenza using improved gaussian process model.*Journal of biomedical informatics*, 93:103144, 2019.
[116]J. P. Linge, R. Steinberger, T. P. Weber, R. Yangarber, E. van der Goot, D. H. Al Khudhairy, and N. I. Stilianakis.Internet surveillance systems for early alerting of health threats.*Eurosurveillance*, 14(13):19162, April 2009.Publisher: European Centre for Disease Prevention and Control. [117]R. Kaiser and D. Coulombier.Different approaches to gathering epidemic intelligence in Europe.*Weekly releases (1997–2007)*, 11(17):2948, April 2006.Publisher: European Centre for Disease Prevention and Control.

[118]C. Paquet, D. Coulombier, R. Kaiser, and M. Ciotti.Epidemic intelligence: a new framework for strengthening disease surveillance in Europe. *Eurosurveillance*, 11(12):5–6, December 2006.Publisher: European Centre for Disease Prevention and Control.
[119]D Coulombier, A Pinto, and M Valenciano.[Epidemiological surveillance during humanitarian emergencies].*Med Trop (Mars)*, 62(4):391–395, January 2002.
[120]The Epidemic Intelligence from Open Sources Initiative.The Epidemic Intelligence from Open Sources Initiative.
[121]Philanthropy Programs for Underserved Communities Google.org.Philanthropy Programs for Underserved Communities - Google.org. [122]Twitter Crunchbase Company Profile & Funding.Twitter - Crunchbase Company Profile & Funding.
[123]Vinay Kumar Jain and Shishir Kumar.An Effective Approach to Track Levels of Influenza-A (H1N1) Pandemic in India Using Twitter.*Procedia Computer Science*, 70:801–807, 2015.
[124]Ali Alessa and Miad Faezipour.Preliminary Flu Outbreak Prediction Using Twitter Posts Classification and Linear Regression With Historical Centers for Disease Control and Prevention Reports: Prediction Framework Study.*JMIR Public Health and Surveillance*, 5(2):e12383, June 2019.
Company: JMIR Public Health and Surveillance Distributor: JMIR Public Health and Surveillance Institution: JMIR Public Health and Surveillance Label: JMIR Public Health and Surveillance Publisher: JMIR Publications Inc., Toronto, Canada.
[125]Aditya Joshi, Ross Sparks, Sarvnaz Karimi, Sheng-Lun Jason Yan, Abrar Ahmad Chughtai, Cecile Paris, and C. Raina MacIntyre.Automated monitoring of tweets for early detection of the 2014 Ebola epidemic.*PLoS ONE*, 15(3):e0230322, March 2020.
[126]Nuha Noha Fakhry, Evan Asfoura, and Gamal Kassam. Tracking Coronavirus Pandemic Diseases using Social Media: A Machine Learning Approach. *IJACSA*, 11(10), 2020.
[127]Samina Amin, Muhammad Irfan Uddin, Duaa H. alSaeed, Atif Khan, and Muhammad Adnan.Early Detection of Seasonal Outbreaks from Twitter Data Using Machine Learning Approaches.*Complexity*, 2021:e5520366, March 2021.Publisher: Hindawi.
[128]Felix A. Gers, Douglas Eck, and Jürgen Schmidhuber.Applying LSTM to Time Series Predictable Through Time-Window Approaches.In Roberto Tagliaferri and Maria Marinaro, editors, *Neural Nets WIRN Vietri-01*, Perspectives in Neural Computing, pages 193–200, London, 2002. Springer. [129]AhmetKara.Multi-stepinfluenzaoutbreakforecastingusingdeepLSTMnetworkandgeneticalgorithm.*ExpertSystemswithApplications*, 180:115153, October 2021.
[130]Kwangok Lee, Munkyu Lee, and Inseop Na.Predicting Regional Outbreaks of Hepatitis A Using 3D LSTM and Open Data in Korea.*Electronics*, 10(21):2668, October 2021.
[131]Nurul Absar, Nazim Uddin, Mayeen Uddin Khandaker, and Habib Ullah.The efficacy of deep learning based LSTM model in forecasting the outbreak of contagious diseases.*Infectious Disease Modelling*, 7(1):170–183, March 2022.
[132]Wenxiao Jia, Xiang Li, Kewei Tan, and Guotong Xie.Predicting the outbreak of the hand-foot-mouth diseases in China using recurrent neural network.In *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 1–4, June 2019.ISSN: 2575-2634.
[133]Sourabh Shastri, Kuljeet Singh, Astha Sharma, Mohamed Lounis, Sachin Kumar, and Vibhakar Mansotra.Chapter 21 - Convolutional bi-directional long-short-term-memory based model to forecast COVID-19 in Algeria.In Rajeev Agrawal, M. A. Ansari, R. S. Anand, Sweta Sneha, and Rajat Mehrotra, editors, *Computational Intelligence in Healthcare Applications*, pages 331–343. Academic Press, January 2022.
[134]Sourabh Shastri, Kuljeet Singh, Sachin Kumar, Paramjit Kour, and Vibhakar Mansotra.Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study.*Chaos, Solitons & Fractals*, 140:110227, November 2020.
[135]Parul Arora, Himanshu Kumar, and Bijaya Ketan Panigrahi.Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India.*Chaos, Solitons & Fractals*, 139:110017, October 2020.
[136]PM Arunkumar, Lakshmana Kumar Ramasamy, et al.Time-series forecasting and analysis of covid-19 outbreak in highly populated countries: A data-driven approach.*International Journal of E-Health and Medical Communications (IJEHMC)*, 13(2):1–17, 2021.
[137]Joshua D. Zelek, John S. Zelek, and Alexander Wong.Why Can't Neural Networks Forecast Pandemics Better.*Journal of Computational Vision and Imaging Systems*, 6(1):1–5, 2020.Number: 1.
[138]Sakinat Oluwabukonla Folorunso, Joseph Bamidele Awotunde, Oluwatobi Oluwaseyi Banjo, Ezekiel Adebayo Ogundepo, and Nureni Olawale Adeboye.Comparison of Active COVID-19 Cases per Population Using Time-Series Models.*IJEHMC*, 13(2):1–21, July 2021.Publisher: IGI Global. [139]Vasilis Papastefanopoulos, Pantelis Linardatos, and Sotiris Kotsiantis.COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population.*Applied Sciences*, 10(11):3880, January 2020.Number: 11 Publisher: Multidisciplinary Digital Publishing Institute. [140]Weiqiu Jin, Shuqing Dong, Chengqing Yu, and Qingquan Luo.A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning.*Computers in Biology and Medicine*, 146:105560, July 2022.
[141]Yuehan Ai, Fan He, Emma Lancaster, and Jiyoung Lee.Application of machine learning for multi-community COVID-19 outbreak predictions with wastewater surveillance.*PLOS ONE*, 17(11):e0277154, November 2022.Publisher: Public Library of Science.
[142]Daren Zhao, Ruihua Zhang, Huiwu Zhang, and Sizhang He.Prediction of global omicron pandemic using ARIMA, MLR, and Prophet models.*Sci Rep*, 12(1):18138, October 2022.Number: 1 Publisher: Nature Publishing Group.
[143]Christophorus Beneditto Aditya Satrio, William Darmawan, Bellatasya Unrica Nadia, and Novita Hanafiah.Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET.*Procedia Computer Science*, 179:524–532, January 2021.
[144]Sujata Dash, Chinmay Chakraborty, Sourav K. Giri, and Subhendu Kumar Pani.Intelligent computing on time-series data analysis and prediction of COVID-19 pandemics.*Pattern Recognition Letters*, 151:69–75, November 2021.
[145]Gopi Battineni, Nalini Chintalapudi, and Francesco Amenta.Forecasting of COVID-19 epidemic size in four high hitting nations (USA, Brazil, India and Russia) by Fb-Prophet machine learning model.*Applied Computing and Informatics*, ahead-of-print(ahead-of-print), January 2020.

[146]Mohammed Ali Shaik and Dhanraj Verma.Deep learning time series to forecast COVID-19 active cases in INDIA: a comparative study.*IOP Conf. Ser.: Mater. Sci. Eng.*, 981(2):022041, December 2020.Publisher: IOP Publishing.
[147]Muzaffer Balaban.Growth Models for Covid-19 Death Figures of Turkey.*Journal of Advances in Medicine and Medical Research*, 32:1–11, November 2020.
[148]Sujata Dash, Chinmay Chakraborty, Sourav Kumar Giri, Subhendu Kumar Pani, and Jaroslav Frnda.BIFM: Big-Data Driven Intelligent Forecasting Model for COVID-19.*IEEE Access*, 9:97505–97517, 2021.Conference Name: IEEE Access.
[149]Lingling Zhou, Ping Zhao, Dongdong Wu, Cheng Cheng, and Hao Huang.Time series model for forecasting the number of new admission inpatients.*BMC Medical Informatics and Decision Making*, 18(1):39, June 2018.
[150]Lorena Saliaj and Eugenia Nissi.Artificial Neural Networks for COVID-19 Time Series Forecasting.*Open Journal of Statistics*, 12(2):277–290, March 2022.Number: 2 Publisher: Scientific Research Publishing.
[151]İsmail Kırbaş, Adnan Sözen, Azim Doğuş Tuncer, and Fikret Şinasi Kazancıoğlu.Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches.*Chaos, Solitons & Fractals*, 138:110015, September 2020.
[152]Eric J. Topol.High-performance medicine: the convergence of human and artificial intelligence.*Nat Med*, 25(1):44–56, January 2019.
[153]Nenad Tomašev, Xavier Glorot, Jack W. Rae, Michal Zielinski, Harry Askham, Andre Saraiva, Anne Mottram, Clemens Meyer, Suman Ravuri, Ivan Protsyuk, Alistair Connell, Cían O. Hughes, Alan Karthikesalingam, Julien Cornebise, Hugh Montgomery, Geraint Rees, Chris Laing, Clifton R. Baker, Kelly Peterson, Ruth Reeves, Demis Hassabis, Dominic King, Mustafa Suleyman, Trevor Back, Christopher Nielson, Joseph R. Ledsam, and Shakir Mohamed.A Clinically Applicable Approach to Continuous Prediction of Future Acute Kidney Injury.*Nature*, 572(7767):116–119, August 2019.
[154]Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng, and Bo Xu.A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19).*Eur Radiol*, 31(8):6096–6104, August 2020.
[155]Chuansheng Zheng, Xianbo Deng, Qiang Fu, Qiang Zhou, Jiapei Feng, Hui Ma, Wenyu Liu, and Xinggang Wang.Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label, March 2020.Pages: 2020.03.12.20027185.
[156]Farah Shahid, Aneela Zameer, and Muhammad Muneeb.Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM.*Chaos, Solitons & Fractals*, 140:110212, November 2020.
[157]Ahmed Ben Said.Predicting COVID-19 cases using bidirectional LSTM on multivariate time series.*Environ Sci Pollut Res*, page 10, 2021. [158]Junling Luo, Zhongliang Zhang, Yao Fu, and Feng Rao.Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms.*Results in Physics*, 27:104462, August 2021.
[159]Hossein Abbasimehr, Reza Paki, and Aram Bahrini.A novel approach based on combining deep learning models with statistical methods for COVID-19 time series forecasting.*Neural Comput & Applic*, 34(4):3135–3149, February 2022.
[160]Hongru Du, Ensheng Dong, Hamada S Badr, Mary E Petrone, Nathan D Grubaugh, and Lauren M Gardner.A deep learning approach to forecast short-term covid-19 cases and deaths in the us.*medRxiv*, pages 2022–08, 2022.
[161]Novel Corona Virus 2019 Dataset.Novel Corona Virus 2019 Dataset.
[162]Population by Country 2020.Population by Country - 2020.
[163]Sangwon Chae, Sungjun Kwon,and Donghyun Lee.Predictinginfectious disease usingdeeplearning and bigdata.*International journalof environmental research and public health*, 15(8):1596, 2018.
[164]Zifeng Yang, Zhiqi Zeng, Ke Wang, Sook-San Wong, Wenhua Liang, Mark Zanin, Peng Liu, Xudong Cao, Zhongqiang Gao, Zhitong Mai, Jingyi Liang, Xiaoqing Liu, Shiyue Li, Yimin Li, Feng Ye, Weijie Guan, Yifan Yang, Fei Li, Shengmei Luo, Yuqi Xie, Bin Liu, Zhoulang Wang, Shaobo Zhang, Yaonan Wang, Nanshan Zhong, and Jianxing He.Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions.*J Thorac Dis*, 12(3):165–174, March 2020.
[165]Juhyeon Kim and Insung Ahn.Infectious disease outbreak prediction using media articles with machine learning models.*Sci Rep*, 11(1):4413, February 2021.Number: 1 Publisher: Nature Publishing Group.
[166]Elham Afzali, Adeola Adegoke, Zhiyong Jin, Woming Qiu, and Liqun Wang.Hybrid VAR-LSTM Networks Modeling and Forecasting COVID-19 Data in Canada.page 11, 2020.
[167]Shwet Ketu and Pramod Kumar Mishra.India perspective: CNN-LSTM hybrid deep learning model-based COVID-19 prediction and current status of medical resource availability.*Soft Comput*, 26(2):645–664, January 2022.
[168]Ankan Ghosh Dastider, Farhan Sadik, and Shaikh Anowarul Fattah.An integrated autoencoder-based hybrid CNN-LSTM model for COVID-19 severity prediction from lung ultrasound.*Computers in Biology and Medicine*, 132:104296, May 2021.
[169]Zuhaira M. Zain and Nazik M. Alturki.COVID-19 Pandemic Forecasting Using CNN-LSTM: A Hybrid Approach.*Journal of Control Science and Engineering*, 2021:1–23, July 2021.
[170]Abdelkader Dairi, Fouzi Harrou, Abdelhafid Zeroual, Mohamad Mazen Hittawe, and Ying Sun.Comparative study of machine learning methods for COVID-19 transmission forecasting.*Journal of Biomedical Informatics*, 118:103791, June 2021.
[171]Sitanath Biswas and Sujata Dash.LSTM-CNN Deep Learning–Based Hybrid System for Real-Time COVID-19 Data Analysis and Prediction Using Twitter Data.In Subhendu Kumar Pani, Sujata Dash, Wellington P. dos Santos, Syed Ahmad Chan Bukhari, and Francesco Flammini, editors, *Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis*, pages 239–257. Springer International Publishing, Cham, 2022.

[172]L. J. Muhammad, Ahmed Abba Haruna, Usman Sani Sharif, and Mohammed Bappah Mohammed.CNN-LSTM deep learning based forecasting model for COVID-19 infection cases in Nigeria, South Africa and Botswana.*Health Technol.*, 12(6):1259–1276, November 2022.
[173]SUNXiao,YEJiaqi,andRENFuji.Detectinginfluenzastatesbasedonhybridmodelwithpersonalemotionalfactorsfromsocialnetworks. *Neurocomputing*, 210:257–268, 2016.
[174]Xiao Sun, Fuji Ren, and Jiaqi Ye.Trends detection of flu based on ensemble models with emotional factors from social networks.*IEEJ Transactions on Electrical and Electronic Engineering*, 12(3):388–396, 2017._eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/tee.22389.
[175]Xiaowei Xu, Xiangao Jiang, Chunlian Ma, Peng Du, Xukun Li, Shuangzhi Lv, Liang Yu, Qin Ni, Yanfei Chen, Junwei Su, Guanjing Lang, Yongtao Li, Hong Zhao, Jun Liu, Kaijin Xu, Lingxiang Ruan, Jifang Sheng, Yunqing Qiu, Wei Wu, Tingbo Liang, and Lanjuan Li.A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia.*Engineering*, 6(10):1122–1129, October 2020.
[176]Denis A. Pustokhin, Irina V. Pustokhina, Phuoc Nguyen Dinh, Son Van Phan, Gia Nhu Nguyen, Gyanendra Prasad Joshi, and Shankar K.An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19.*Journal of Applied Statistics*, pages 1–18, November 2020.
[177]Evan L Ray, Nutcha Wattanachit, Jarad Niemi, Abdul Hannan Kanji, Katie House, Estee Y Cramer, Johannes Bracher, Andrew Zheng, Teresa K Yamana, Xinyue Xiong, et al.Ensemble forecasts of coronavirus disease 2019 (covid-19) in the us.*MedRXiv*, pages 2020–08, 2020.
[178]Sina F Ardabili, Amir Mosavi, Pedram Ghamisi, Filip Ferdinand, Annamaria R Varkonyi-Koczy, Uwe Reuter, Timon Rabczuk, and Peter M Atkinson. Covid-19 outbreak prediction with machine learning.*Algorithms*, 13(10):249, 2020.
[179]Sweeti Sah, B. Surendiran, R. Dhanalakshmi, Sachi Nandan Mohanty, Fayadh Alenezi, and Kemal Polat.Forecasting COVID-19 Pandemic Using Prophet, ARIMA, and Hybrid Stacked LSTM-GRU Models in India.*Computational and Mathematical Methods in Medicine*, 2022:1–19, May 2022. [180]Kehua Guo, Changchun Shen, Xiaokang Zhou, Sheng Ren, Min Hu, Minxue Shen, Xiang Chen, and Haifu Guo.Traffic Data-Empowered XGBoost-LSTM Framework for Infectious Disease Prediction.*IEEE Transactions on Intelligent Transportation Systems*, pages 1–12, 2022.Conference Name: IEEE Transactions on Intelligent Transportation Systems.
[181]Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani.Multiscale mobility networks and the spatial spreading of infectious diseases.*Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.
[182]Vittoria Colizza, Alain Barrat, Marc Barthelemy, Alain-Jacques Valleron, and Alessandro Vespignani.Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions.*PLoS medicine*, 4(1):e13, 2007.
[183]Duygu Balcan, Hao Hu, Bruno Goncalves, Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, Wouter Van den Broeck, et al.Seasonal transmission potential and activity peaks of the new influenza a (h1n1): a monte carlo likelihood analysis based on human mobility.*BMC medicine*, 7(1):1–12, 2009.
[184]Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang.Modelling disease outbreaks in realistic urban social networks.*Nature*, 429(6988):180–184, 2004.
[185]Neil M Ferguson, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke.Strategies for mitigating an influenza pandemic.*Nature*, 442(7101):448–452, 2006.
[186]Joshua M Epstein, D Michael Goedecke, Feng Yu, Robert J Morris, Diane K Wagener, and Georgiy V Bobashev.Controlling pandemic flu: the value of international air travel restrictions.*PloS one*, 2(5):e401, 2007.
[187]Marta Luisa Ciofi degli Atti, Stefano Merler, Caterina Rizzo, Marco Ajelli, Marco Massari, Piero Manfredi, Cesare Furlanello, Gianpaolo Scalia Tomba, and Mimmo Iannelli.Mitigation measures for pandemic influenza in italy: an individual based model considering different scenarios.*PloS one*, 3(3):e1790, 2008.
[188]Florian Rohart, Gabriel J Milinovich, Simon MR Avril, Kim-Anh Lê Cao, Shilu Tong, and Wenbiao Hu.Disease surveillance based on internet-based linear models: an australian case study of previously unmodeled infection diseases.*Scientific reports*, 6(1):1–11, 2016.
[189]Sungjin Cho, Chang Hwan Sohn, Min Woo Jo, Soo-Yong Shin, Jae Ho Lee, Seoung Mok Ryoo, Won Young Kim, and Dong-Woo Seo.Correlation between national influenza surveillance data and google trends in south korea.*PloS one*, 8(12):e81422, 2013.
[190]Yue Teng, Dehua Bi, Guigang Xie, Yuan Jin, Yong Huang, Baihan Lin, Xiaoping An, Dan Feng, and Yigang Tong.Dynamic forecasting of zika epidemics using google trends.*PloS one*, 12(1):e0165085, 2017.
[191]Andrea Freyer Dugas, Mehdi Jalalpour, Yulia Gel, Scott Levin, Fred Torcaso, Takeru Igusa, and Richard E Rothman.Influenza forecasting with google flu trends.*PloS one*, 8(2):e56176, 2013.
[192]Nimai Chand Das Adhikari, Arpana Alka, Vamshi Kumar Kurva, Suhas S, Hitesh Nayak, Kumar Rishav, Ashish Kumar Nayak, Sankalp Kumar Nayak, Vaisakh Shaj, and Karthikeyan.Epidemic Outbreak Prediction Using Artificial Intelligence.*IJCSIT*, 10(4):49–64, August 2018.
[193]Prosper Yeng, Ashenafi Zebene Woldaregay, and Gunnar Hartvigsen.K-cusum: Cluster detection mechanism in edmon.2019.
[194]Alvin Wei Ze Chew, Yue Pan, Ying Wang, and Limao Zhang.Hybrid deep learning of social media big data for predicting the evolution of covid-19 transmission.*Knowledge-Based Systems*, 233:107417, 2021.
[195]Giles Hogben.Security issues and recommendations for online social networks.*ENISA position paper*, 1:1–36, 2007.
[196]Marcel Salathé.Digital epidemiology: what is it, and where is it going?page 6, 2018.
[197]Joseph S Lombardo, Joel C Gaydos, et al.A public health role for internet search engine query data?*Military medicine*, 174(8):XI, 2009. [198]Gabriel J Milinovich, Gail M Williams, Archie CA Clements, and Wenbiao Hu.Internet-based surveillance systems for monitoring emerging infectious diseases.*The Lancet infectious diseases*, 14(2):160–168, 2014.