

A systematic literature review of time series methods applied to epidemic prediction

Apollinaire Batoure Bamana ^{a,*}, Mahdi Shafiee Kamalabad ^{b,c}, Daniel L. Oberski ^{b,d}

^a Department of Computer Engineering, University Institute of Technology, The University of Ngaoundere, Ngaoundere, Cameroon

^b Department of Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, The Netherlands

^c Centre for Complex Systems Studies, Utrecht University, Utrecht, The Netherlands

^d Utrecht University Medical Center, Utrecht University, Utrecht, The Netherlands

ARTICLE INFO

Keywords:

Time series forecasting
Time series methods
Epidemic prediction
Missing data
Outlier data
African context
COVID-19 impact

ABSTRACT

While time series are extensively utilized in economics, finance and meteorology, their application in epidemics has been comparatively limited. To facilitate a comprehensive research endeavor on this matter, we deemed it necessary to commence with a systematic literature review (SLR). This Systematic Literature Review aims to assess, based on a sample of relevant papers, the use of Time Series Methods (TSM) in epidemic prediction, with a special focus on African issues and the impact of COVID-19. The SLR was conducted using databases such as ACM, IEEE, PubMed and Science Direct. Open access published papers in English, in a peer reviewed Journals, from 2014 to 2023, containing keywords such as Time Series, Epidemic and Prediction were selected. The findings were summarized in an adapted PRISMA flow diagram. We end up with a sample of 36 papers. As conclusion, TSM are not so used in epidemic prediction as in some other domains, even though epidemic data are collected as time series. Just very few works address African issues regarding diseases and countries. COVID-19 is the pandemic that revealed and enhanced the used of TSM to forecast epidemics. This work paves ways for R&D on epidemiology, based on TSM.

Contents

1. Introduction	2
2. Methodology	2
2.1. Objectives, search terms and equation	2
2.2. Inclusion/exclusion and screening criteria	2
2.3. Analysis of selected papers	2
3. Results	3
3.1. Overview and synthesis of selected articles	3
3.2. Assessing objectives set	4
4. Discussion	7
5. Conclusion	11
CRediT authorship contribution statement	11
Declaration of competing interest	11
Acknowledgments	12
References	12

* Corresponding author.

E-mail address: apollinaire.batoure@univ-ndere.cm (A. Batoure Bamana).

1. Introduction

The significance of time series analysis has become paramount across various domains, owing to its diverse range of applications. In this paper, we aim to harness these analytical methods to forecast epidemics, with a particular focus on African contexts, and the COVID-19 impact.

A time series refers to a sequence of events observed and recorded over a period of time [1]. The collected data is subjected to analysis, enabling predictions and providing decision-making support [2,3]. Time series find applications in various disciplines such as astronomy, meteorology, seismology, signal theory, biology, geophysics, economics, stock exchange, demography, and industry [4,5]. Ongoing research and development efforts in this domain indicate its active pursuit [1]. Time series data is ubiquitous, as time is an inherent component of everything that is observable [6]. Methods utilized for analyzing time series make effective use of statistics, probability, and in contemporary times, Machine Learning (ML) and Deep Learning (DL) [7–13].

Time series applications play a crucial role in public health, encompassing a range of functions for the effective management of epidemics. The Centers for Disease Control and Prevention (CDC¹) describes an epidemic or outbreak as an unexpected increase in the number of disease cases in a specific geographical area. An outbreak is characterized by its sudden onset, rapid spread, and exponential growth or development.² Efficient anticipation is essential to minimize casualties during outbreaks, as it enables the implementation of timely and appropriate responses [14]. By anticipating the potential challenges and risks, proactive measures can be taken to effectively manage the population health [15–17].

Some relevant SLR were conducted on ML, DL or AI, for epidemic prediction. We can state the works in [18–22]. All the SLR articles up cited focus on methods using time series data. [21] encompasses 16 articles within the sample of 64 that specifically address spatiotemporal considerations. In the case of [20], 25 out of the 75 selected papers pertain to time series methods. Notably, the predominant method across these studies is the PRISMA,³ employed in three out of the five studies [19–21]. The remaining two studies, namely [18,22], do not adhere to any specific known methodological approach. On the other hand, [20] uses a combination of Cochrane, MOOSE,⁴ and PRISMA, while [21] adopts CHARMS⁵ and PRISMA, with the additional step of registering the review in PROSPERO.⁶ None of these studies incorporates the SPIDER⁷ method, which is particularly useful in formulating research questions. To the best of our knowledge, there is currently no existing systematic literature review focusing on the prediction of epidemics and cases, specifically employing time series methods. The above issue is the cornerstone of this work's contribution.

In next sections, we first state the research methodology employed and then present and analyze the obtained articles sample. We finally engage in a comprehensive discussion and concluded the work.

2. Methodology

In conducting the SLR on employing time-series methods for epidemic prediction, we implemented a tailored methodology, delineated into specific components outlined in the subsequent sections.

¹ <https://www.cdc.gov/>

² <https://www.merriamwebster.com/dictionary/epidemic>

³ Preferred Reporting Items for Systematic Reviews and Meta-Analyses

⁴ Meta-Analysis Of Observational Studies in Epidemiology

⁵ Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies

⁶ International Prospective Register of Systematic Reviews (<https://www.crd.york.ac.uk/prospero/>)

⁷ Sample, Phenomenon of interest, Design, Evaluation, Research type

2.1. Objectives, search terms and equation

The objectives of this SLR are threefold:

- Comprehensive evaluation of use of TSM in epidemic prediction;
- Focused assessment of TSM in the African context;
- Analysis of the impact of COVID-19 on research, using TSM.

These objectives collectively aim to offer a comprehensive understanding of the role and impact of TSM in epidemic prediction, with specific attention given to diseases occurring in African Countries, and the transformative effects brought by the COVID-19 pandemic.

We established the research question based on the SPIDER tool [23] formulated as: *How time series help predicting epidemics occurring in human population, and number of cases?*

Defining research question opens way for search terms and equation. Search terms are keywords that help querying databases [21,22]. The selected search terms and their synonyms are:

- *Time Series*
- *Epidemic / Outbreak / Disease*
- *Prediction / Forecasting / Detection / Surveillance.*

All search terms or their synonyms must be present in the document's title, keywords, or abstract when querying repositories. Evaluating this requirement results in the formulation of a general search equation using the AND with OR logic operator as follows: *Time Series AND (Epidemic OR Outbreak OR Disease) AND (Forecasting OR Prediction OR Detection OR Surveillance)*. Subsequently, the search equation is adapted according to each repository's specificity.

We selected four repositories to retrieve scientific works. They are PubMed (pubmed.ncbi.nlm.nih.gov/), ACM (dl.acm.org/), IEEE Xplore (ieeexplore.ieee.org/) and Science Direct (sciencedirect.com/).

2.2. Inclusion/exclusion and screening criteria

Inclusion/Exclusion criteria are open access papers, written in English and published in a peer reviewed journal, between 2014 and 2023 (10 years). The paper should not be a review, meta-analysis paper, trial, preprint, book chapter, conference presentation, opinions, commentary or dissertation.

The screening stage brings out relevant documents based on their content. The paper screening process here involves sequential assessments based on title, keywords, abstract, and methodology in each preselected paper. The screening criteria encompasses exclusion of papers:

- On non-communicable diseases such as Alzheimer's, Parkinson's, Sepsis, Cardiovascular diseases, Diabetes, Hypertension, etc.;
- On impact of COVID-19 on the activities or on other diseases, unless the disease under consideration is communicable;
- On topics related to animals or plants, vaccines, suicide, smoke, drugs, clinical trials, surveys, etc.;
- On expert systems and mathematical modeling;
- Duplicated or withdrawn;
- Lacking a clear focus on predicting or detecting epidemics or number of cases occurring;
- Not falling within the Computer Science Domain.

2.3. Analysis of selected papers

The analysis criteria defined here includes various dimensions, namely Year of publication of the paper; Country and Continent where the work is carried out; Disease addressed; Time series methods performed, as well as the category of the method; Data source; Journal where the paper is published; Mixed or combined methods performed; Main idea or objective of the work; Main findings; Characteristic of dataset; Evaluation metrics used; Programming language or tool

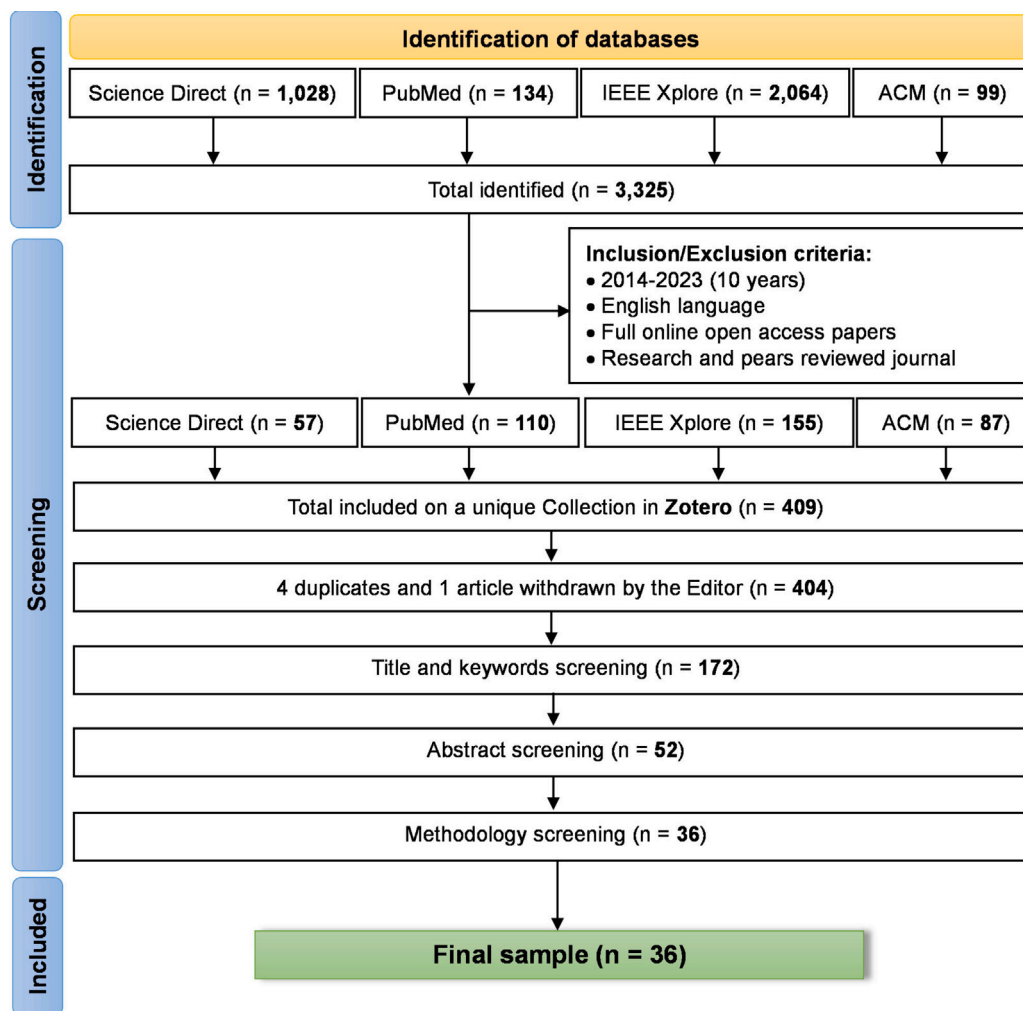


Fig. 1. Flow Diagram of Systematic Search Results (based on PRISMA).

utilized for experimentation; and finally the way missing and outliers data are handled. Analysis is categorized into four distinct groups: Time Series Models, Machine Learning, Deep Learning, and Mixed models.

This methodological framework has proven instrumental in yielding the following results.

3. Results

From the 3,325 references initially collected in repositories, we applied a number of filters based on the inclusion/exclusion criteria. By applying these criteria, we obtained 57 separate references from Science Direct; 110 from PubMed; 155 from IEEE and 87 from ACM. We get a total of 409 preselected references. The results were grouped into a single Collection created in Zotero software [24]. From the initial collection, we removed 4 duplicates and 1 article withdrawn by the Editor, resulting in a reference base of 404 items. Screening based on title and keywords narrowed down the selection to 172 references. Subsequently, abstract screening further refined the selection to 52 shortlisted works. The final screening was based on methodology, resulting in a sample of 36 articles for analysis. This process is summarized in the flow chart depicted in Fig. 1, adapted from a PRISMA⁸ model. It illustrates the databases selected and the initial querying results obtained (*Identification*), the refining operations carried

out (*Screening*), and the final sample considered (*Included*). Table 1 presents the main abbreviations used in the document.

The next sections first unveil an overview and synthesis of the work carried out, using a series of tables; Second assess the objectives set.

3.1. Overview and synthesis of selected articles

A total of 36 articles, meeting all the specified criteria, have been included in our analysis sample. Table 2 provides a summary of these articles, capturing key details such as the Year of publication, Country (or Countries) and Continent where the study was conducted, Methods employed (categorized as TSM, ML, DL, or Mixed), specific Disease addressed, Data source, and Journal of publication. Subsequent to the general information table (Table 2), we have organized a summary for each of the four method categories: TS Models categories (Table 3), ML models (Table 4), DL models (Table 5), and Mixed models (Table 6). Within each category, we highlight the main idea or objective, primary results, dataset characteristics, evaluation metrics, and the language or tool used for experimentation. Building upon these individual summaries, an overarching quantitative summary is presented in Tables 7 to 17. These tables group information based on: method categories, combined methods, metrics considered, languages used, data sources, journals, year of publication, diseases treated, most commonly employed models, countries and continents of study. This extensive dataset, along with information on missing and outlier data, is used to fulfill the objectives outlined in this research work. All the abbreviations used in the above mentioned tables are listed in Table 1.

⁸ Preferred Reporting Items for Systematic Reviews and Meta-Analyses (www.prisma-statement.org/)

Table 1

Abbreviations.

AIC: Akaike Information Criteria	LSTM: Long Short-Term Memory
ANN: Artificial Neural Networks	M5PRT: M5P Regression Tree
APE: Absolute Percentage Error	MAD: Mean Absolute Deviation
ARIMA: Autoregressive Integrated Moving Average	MAPE: Mean Absolute Percentage Error
AUC: Area Under the ROC Curve	MAE: Mean Absolute Error
BIC: Bayesian Information Criteria	MHEM: Maximum-Hasting Estimation Method
BMA: Bayesian Model Averaging	MLP: Multi Layer Perceptron
BR: Bayesian Regression	MSE: Mean Square Error
CDC: Center for Disease Control and prevention	NARNN: Nonlinear Autoregressive Neural Network
CDO: Climate Data Online	N-BEATS: Neural Basis Expansion Analysis for Time Series
CHSC: Chinese Health and Science Center	NNETAR: Neural NETWORK Autoregressive
CNN: Convolution Neural Network	NOAA: National Oceanic and Atmospheric Administration
CUSUM: CUMulative SUM	NPL: Natural Language Processing
DDEM: Data-Driven Estimation Method	NPV: Negative Predictive Value
DL: Deep Learning	NRM: Non-parametric Regression Model
DLD: Department of Livestock and Development	NYT: New York Times
DSPM: Deep Sequential Prediction Model	PC: Pearsons Correlation
DTR: Decision Tree Regressor	PCC: Pearson Correlation Coefficient
EARS: Early Aberration Reporting System	PMD: Peak Magnitude Difference
ENR: Elastic Net Regressor	PPV: Positive Predictive Value
ES: Exponential Smoothing	PWD: Peak Weak Difference
ETS: Error Trend Seasonality	RF: Random Forest
FNN: Feedforward Neural Networks	RFR: Random Forest Regressor
FP: Facebook Prophet	RMSE: Root-Mean-Square Error
GEP: GEnetic Programation	RNN: Recurrent neural network
GFT: Google Flu Trends	SARIMA: Seasonal ARIMA
GRU: Gated Recurrent Units	sMAPE: Symmetric Mean Absolute Percent Error
GTD: Google Trend Data	SP-CPC: Spatial Probabilistic Contrastive Predictive Coding
HDX : Humanitarian Data eXchange	SSA: Sparrow Search Algorithm
HFMD/HFD: Hand-Foot-Mouth Disease	SSM-U/M: Space-State Model Univariate/Multivariate
HWES: Holt-Winters Exponential Smoothing	STRW: Seasonal Trend Random Walk
IDWR: Infectious Diseases Weekly Report	SVR/M: Support Vector Regression/Machine
IEDCR: Institute of Epidemiology Disease Control and Research	TBATS: Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components
ILI : Influenza Like Illness	TCN: Temporal Convolutional Network
JHUC: Johns Hopkins University Center	TR: Theilsen Regression
KNR: K Nearest Regressor	TSFEL: Time Series Feature Extraction Library
KRR: Kernel Ridge Regressor	VAR: Vector Autoregressive
LR: Linear Regression	WEKA: Waikato Environment for Knowledge Analysis

3.2. Assessing objectives set

Objective 1: Comprehensive evaluation of time series methods in epidemic prediction.

Of the 36 articles, the COVID-19 pandemic, which occurred at the end of 2019 in Wuhan, China [61], comes far out on top, with more than two third of the articles. COVID-19 is followed by influenza, which is the focus of four papers [26,34,45,60], Dengue fever with two papers [30,56], HFMD diseases, also with two papers [57,59], and finally diarrhoea and tuberculosis, which each has one paper devoted [29,42].

In terms of publication years, the dominance of COVID-19 is evident, accounting for 26 out of 36 publications. This indicates the major publication years. Over the ten-year period from 2014 to 2023, 12 works were published in 2021 and 8 in 2022. In 2020 and 2023, 6 works were published each. The year 2019 contained 2 publications. Finally, 2018 and 2014 each has one article.

The works are categorized into four main methodological groups: TSM, ML, DL, and combined methods. The most frequently utilized method category is DL, with 12 papers. ML and TSM methods comprise

6 and 3 works, respectively. It is noteworthy that there is a trend towards combining methods, with 15 out of 36 projects employing combined approaches. These combinations include TSM and DL (05 works, [25,43,48,53,59]); TSM and ML (03 works, [42,44,58]; ML and DL (04 works, [33,40,49,50]) and finally, TSM, DL and ML (03 works, [36,46,60]).

Regarding the chosen databases, there are 8 journals in which the works have been published. IEEE and ACM lead with 9 and 8 papers, respectively. Elsevier and Science Direct follow closely with 6 and 5 works each. MDPI, Springer, PLoS ONE, and Wiley have fewer than 5 papers each.

The primary sources of data used are CDC and JHUC, with 7 and 6 studies, respectively, utilizing datasets from these databases. Other commonly used sources include Google and WHO. In total, 14 different data sources were identified across the 36 studies.

In terms of experimental languages or tools used, 23 out of the 36 works mention them. Among these 23 works, 15 utilized the Python scientific programming language and related libraries, while 4 used R

Table 2
General information on papers.

Ist Author	Year	Country	Continent	Methods	Category	Disease	Data source	Journal
Rasjid [25]	2020	Indonesia	Asia	ES, LSTM	Mixed	COVID-19	HDX	Science Direct
Borghi [26]	2021	Brazil, Italy	America, Europe	MLP, ANN	DL	COVID-19	JHUC	Science Direct
Venna [27]	2019	USA	America	LSTM	DL	Influenza	CDC, GFT	IEEE Access
Safari [28]	2021	USA, Brazil, Russia, India, Peru, Spain, Italy, Iran, Germany, U.K.	America, Europe, Asia	LSTM	DL	COVID-19	WHO	Elsevier
Yang [29]	2018	Korea	Asia	CUSUM, EARS, ARIMA, HWES	TSM	Diarrhea infection	KCDC	MDPI
Ferdousi [30]	2021	Brazil	America	RNN	DL	Dengue fever	Info Dengue Project, NOAA	IEEE Access
Nuanchuay [31]	2021	99 countries around the world		LSTM	ML	COVID-19	WHO	ACM
Liapis [32]	2020	15 countries in South and Central Europe	Europe	LR, SVR, M5PRT	ML	COVID-19	CSSEGISand Data/COVID-19	ACM
Lobato [33]	2021	Argentina, Bolivia, Brazil Colombia, Cuba, Guatemala, Haiti, Mexico, Peru	Latin America	SVR, LSTM	Mixed	COVID-19	Our World in Data website	ACM
Watmaha [34]	2021	USA	America	RNN, LSTM, CNN	DL	Influenza	CDC	ACM
Tetteroo [35]	2022	58 countries worldwide		Auto Machine Learning (AutoML)	ML	COVID-19	ECDC	IEEE ACCESS
Assad [36]	2022	Brazil, Italy, USA	America, Europe	ES, ARIMA, VAR, TBATS, SSM-U, SSM-M, MLP, NNETAR	Mixed	COVID-19	Real COVID-19 data	Springer
Masum [37]	2020	Bangladesh	Asia	LSTM	DL	COVID-19	IEDCR	Science Direct
Kim [38]	2023	USA	America	LSTM, NPL	DL	COVID-19	WHO, GTD	PLoS ONE
Carvalho [39]	2021	Brazil, Portugal	America, Europe	ANN	DL	COVID-19	Worldmeter website	Science Direct
Balli [40]	2020	USA, Germany	America, Europe	LR, MLP, RF, SVM	Mixed	COVID-19	WHO	Elsevier
Motavali [41]	2023	Canada	America	N-BEATS	DL	COVID-19	Government of Saskatchewan	IEEE Access
Liang [42]	2020	China	Asia	SARIMA, SVM, BMA	Mixed	Tuberculosis	CHSC, Baidu index	ACM
Swaraj [43]	2021	India	Asia	ARIMA, NARNN	Mixed	COVID-19	JHUC	Elsevier
Lucic [44]	2021	USA	America	K-Means, ARIMA, STRW	Mixed	COVID-19	CDC	IEEE Open Journal
Li [45]	2021	USA, Japan	America, Asia	Transformer models	DL	Influenza	IDWR, CDC	Elsevier
Kumar [46]	2023	India, the United States, Russia, Argentina, Brazil, Colombia, Italy, Turkey, Germany, France	America, Europe	RFR, DTR, KNR, Lasso Regression, LR, BR, TR, KRR, RANSAC Regressor, XG Boost, ENR, FP, Stacked LSTM, Stacked GRU	Mixed	COVID-19	Kaggle COVID dataset	Springer
Irungu [47]	2023	USA	America	KNN, RF, SVM	ML	COVID-19	Health SOS	IEEE Access
Jin [48]	2023	Quebec (Canada), Italy	America, Europe	CNN, LSTM, ARIMA, TCN, SSA	Mixed	COVID-19	JHUC	IEEE Access
Chen [49]	2021	India, Russia, Chile	Asia, America, Europe	LSTM, Bayesian optimization	Mixed	COVID-19	Statistics from countries	ACM

(continued on next page)

Table 2 (continued).

Ayris [50]	2022	Bangladesh, China, Canada, Australia, Guyana, Montenegro, Algeria, Belarus, Chile, New Zealand, Brazil, Turkey, USA, UK, Colombia, France, El Salvador, Cameroon, Nigeria, India, Senegal, Nepal, Costa Rica, Cuba, Iran, Indonesia, Israel, Lesotho	Africa, America, Asia, Europe	DSPM, NRM	Mixed	COVID-19	JHUC	Elsevier
Moloshnikov [51]	2022	Russia	Europe	LSTM	ML	COVID-19	Yandex Data Lens	Science Direct
Jain [52]	2022	China, Italy, Spain, US, Russia, Brazil	America, Asia, Europe	SVM, Polynomial regression, BRR	ML	COVID-19	JHUC	Wiley
Jin [53]	2023	Germany, Japan	Europe, Asia	ARIMA, LSTM	Mixed	COVID-19	JHUC	IEEE Access
Susarla [54]	2022	USA	America	Contrastive Predictive Coding	ML	COVID-19	New York Times	ACM
Salgotra [55]	2020	India	Asia	Genetic programming	ML	COVID-19	India time series datasets	Elsevier
Somboon-sak [56]	2019	Thailand	Asia	SARIMA	TSM	Dengue Fever	Bureau of Epidemiology	ACM
Feng [57]	2014	China	Asia	SARIMA	TSM	HFMD	CDC	PLoS ONE
Zrieq [58]	2022	Saudi Arabia	Asia	ARIMA, Facebook Prophet	Mixed	COVID-19	Saudi Health Ministry	MDPI
Punyaporn-withaya [59]	2022	Thailand	Asia	SARIMA, ETS, NNAR, TBATS	Mixed	FMD	DLD	MDPI
Olukanmi [60]	2021	South Africa	Africa	LSTM, FNN, MLR, EN, SVM, SARIMA	Mixed	ILI	GTD	IEEE Access

Table 3
Time series models.

1st Author	Main Idea/Objective	Main findings	Characteristics of dataset	Metrics	Language
Yang [29]	Comparison of four different temporal outbreak detection algorithms	Holt–Winters showed better performance when the baseline frequency and the dispersion parameter values are both less than 1.5 and 2, respectively	Simulated data on 624 weeks	PPV, NPV, F1, sMAPE, RMSE, MAD	R
Somboon-sak [56]	Find an appropriate model with optimal parameter values, to predict the occurrence of dengue fever in Thailand	The SARIMA model (1,2,2) (1,1,1)12 is the best model for the prediction	Data of dengue patients from January 2014 to December 2018	MAPE, AIC, BIC	X
Feng [57]	Use of multivariate SARIMA associated with meteorological variable	SARIMA (0,1,0) (1,0,0)52 associated with Tav _g at lag 2 weeks, SARIMA (0,1,2) (1,0,0)52 with Tav _g -Lag 2 weeks and SARIMA (0,1,1) (1,1,0)52 with Tav _g -Lag 3 weeks are proposed for description and predication the weekly number of HFMD	2932 cases with HFMD from Jan. 2008 to Jun. 2012. Atmospheric temperature routinely measured within the same period, and aggregated on a weekly basis with a total of 234 weeks period	RMSE, BIC, R ²	X

and 2 used MatLab. Weka and AutoML were each employed in one paper for experimentation.

Data from 44 countries were considered in the various studies. The most represented countries (with at least 3 studies) are the USA with 13 studies, followed by Brazil with 9 studies. Russia and Italy each have 7 papers, and India has 6. China and Germany both have 4 works, while

Colombia and Canada each have 3 works. 4 continents are represented: America with 19 papers, Asia and Europe with 16 and 13 papers, respectively, while Africa has only two devoted works.

Out of 24 different methods used for time series prediction, 16 are employed at least twice. The LSTM and ARIMA methods lead the way with 14 and 9 papers, respectively. The Support vector, Exponential

Table 4
Machine learning models.

1st Author	Main Idea/Objective	Main findings	Characteristics of dataset	Metrics	Language
Liapis [32]	Combining ensemble method and three machine learning techniques to reduce the forecasting errors	The proposed model outperforms both its base learners and a number of widely-used individual algorithms	30 datasets contained univariate time series data related to the COVID-19	RMSE, MAPE	WEKA
Tetteroo [35]	Adaptation of the AutoML framework to COVID-19 forecasting	Combining mobility features and mortality data improves forecasting accuracy	Data from December 31st, 2019 until December 14th, 2020	RMSE	AutoML frameworks
Irungu [47]	Investigates patients' electro cardiography properties for an accurate prediction of COVID-19	Classification encompassing Temporal, Spectral, and Wavelet features with Random Forest, has a slight edge over other models	Dataset from TSFEL and split into train (75%) and tests (25%) sets respectively	F1 score	Python
Jain [52]	Analysis, prediction, and visualization of COVID-19 data	Results are 99% accurate	Daily confirmed cases worldwide, from January 2020	MEA, MSE, R^2	Python
Susarla [54]	Including human mobility data by using SP-CPC which Contrastive Predictive Coding (CPC)	The model applied to COVID-19 data make accurate short-term predictions, more accurate than ARIMA, one day into the future	Three years data (2020, 2021, 2022), associated to mobility data from the University of Wisconsin	MAPE	X
Salgotra [55]	Robust and reliable variant of GEP to model the confirmed and death cases of COVID-19 in India	The proposed GEP-based models use simple linkage functions and are highly reliable for time series prediction of COVID-19 cases	Total number of confirmed cases and death cases is around 80,000 and 2500 respectively	MSE, RMSE, R^2	X

Smoothing, and SARIMA methods follow, each with 7, 5, and 5 works, respectively.

15 different methods are utilized to assess prediction accuracy. RMSE is the most commonly used, appearing 24 times. Next, we have MAE and MAPE, each used 14 times. R^2 and MSE are performed 6 and 5 times respectively, while the remaining methods are used less than 5 times.

A cross-analysis of the method categories by year (Table 18) revealed that all studies using only TS models were published before 2020. All other methods (ML, DL, and Mixed) were employed in works published between 2020 and 2023, mainly focusing on the COVID-19 pandemic.

Handling missing and outlier data. From our sample of 36 articles, 10 [27,33,35,45,46,48,54,57,58,60] manage missing data, and 5 [33, 46,48,53,60] outliers. Among these works, 04 [33,46,48,60] deal with both missing and outlier data. 6 papers [27,35,45,54,57,58] consider only missing data and 01 [53], only outlier data.

[27] uses Simple moving average-based smoothing, to manage missing data. In his work, [35] imputed missing values by taking the average of the values 7 days before the missing data point, and 7 days after the missing data point. [33,53] simply remove the missing data points before forecasting. [46] identifies missing data, but does not say how they are handled. [48] performs a linear imputation of MDs while [60] uses `tsclean` from the R forecast package to estimate the value of missing data. [54] uses the master FIPS⁹ codes list to fill in missing data. [58] states that the model is quite robust to missing values and [57] reports that the dataset used does not have any missing values.

In terms of outliers, [25] identifies them without specifying the way they are handled. In the work of [46], the management of outliers depends on the method used. According to this work [46], RANSAC¹⁰ automatically excludes outliers; Theilsen regression and Prophet are respectively insensitive and resistant to outliers. [48] uses boxplots to identify outliers and then removes them from the dataset. [60] uses

`tsclean` from R forecast package to estimate replacement values for outliers in the dataset.

Objective 2: Focused assessment of time series methods in the African context

From the above, the increased use of TS methods for epidemic prediction is very recent (after 2019). Among the elements of our sample, only one work focuses specifically on an African country. This is the work of Olukunmi [60], conducted in 2021 on South Africa, published in IEEE and focusing on Influenza. In the work of Ayris [50], carried out in 2022 and published by Elsevier, focusing on COVID-19, data from Algeria, Cameroon, Nigeria and Senegal were used. These two studies, out of the 36, are the only ones to have considered data from African countries.

Objective 3: Analysis of the Impact of Covid-19 on use of time series methods

Among the 36 works included in our sample, 26 are dedicated to the study of COVID-19, all of which were published after 2019. A detailed examination of COVID-19 publications by year is presented in Table 19. Notably, there were 8 studies on COVID-19 conducted in 2021, followed by 7 in 2022, 6 in 2023, and 5 in 2020.

After presenting the results, we now move to discussion.

4. Discussion

Initiating scientific endeavors to leverage time series techniques for predicting epidemics, particularly in African contexts, prompted the necessity for conducting a Systematic Literature Review (SLR). This review encompassed a comprehensive evaluation of time series methods (TSMs) implementation for epidemic prediction over a decade, spanning from 2014 to 2023, and examined the impact of the COVID-19 pandemic on this research domain.

The investigation reveals a notable disparity in the utilization of time series for epidemic forecasting compared to other fields like econometrics, financial markets, meteorology, industry, or transport. Prior to the emergence of COVID-19, only four works [27,29,56,57] within our sample addressed this area, with only one employing a Deep Learning (DL) approach. The remaining three [29,56,57] studies

⁹ Federal Information Processing Standards

¹⁰ RANdom SAmple Consensus algorithm

Table 5
Deep learning models.

1st Author	Main Idea/Objective	Main findings	Characteristics of dataset	Metrics	Language
Borghgi [26]	Describes and predicts the number of infected cases and deaths of COVID-19	The model describes and predicts for up to six days, the behavior of time series related to the infected cases and deaths of COVID19	30 countries with the largest number of confirmed cases selected until 11 may of 2020	MSE, MAE	MatLab
Venna [27]	Introduction of LSTM method to capture the temporal dynamics of seasonal flu and external variables	Proposed method, combining spatio-temporal and environmental factors, performs better than the existing influenza forecasting methods	14 years of flu data from CDC (2002–2016) and 11 years of data from GFT (2003–2014)	MAPE, RMSE	R
Safari [28]	Prediction of the COVID-19 incidence, including new cases, recovery cases, and mortality rate	Model shows robustness and proficiency in uncertain, chaotic and complex time series problems such as the COVID-19 pandemic	Data on number of confirmed, death, and recovery cases during the COVID-19 pandemic	AUC, RMSE	X
Ferdousi [30]	Accurately prediction of dengue fever cases	The method can be generalized and used for projecting any infectious outbreak where temporal data at a reasonable spatial resolution are available	Weekly dengue fever case data on 700 municipalities of Brazil from 2010, associated to weather observation data from NOAA	MAE	Python
Nuanchuay [31]	Use of features of a country and boundaries countries, to improve prediction accuracy of COVID-19	Additional features fed to LSTM model improve predicted results up to 5% better than the LSTM model without additional features	Case report from 237 countries from January 3, 2020, to May 6, 2021	MAE, MAPE	Python
Wtmaha [34]	The research compared models on RNN, LSTM and CNN	CNN has significant influenza forecasting, more effectively than RNN and LSTM	10 regions flu count data of United States, between 1997–2016 from CDC and NCIRD	MAPE, RMSE	Python
Masum [37]	Apply LSTM and compare results with Random Forest or SVR methods	LSTM based is a well-done formula, comparing to Random Forest or Support Vector Regression	Daily report confirmed and recovery cases from March 8 to May 12, collected from IEDCR	RMSE	X
Kim [38]	Predicting COVID-19 outbreaks in five states in the USA using based on LSTM and Seq2Seq+ Attention models	Combination of Word2Vec, LSTM and Seq2Seq + Attention achieves better prediction results than the LSTM and Seq2Seq + Attention models	Data on COVID-19 collected from April 5, 2020, to June 27, 2021; and 49,456 NYT articles written in the same period	RMSE	X
Carvalho [39]	Prediction model with two different approaches, by using ANN	Twenty days ahead forecasting made with an error of 24,7% and 1,6% for the number of cumulative cases of infection and deaths for Brazil, and 37,9% and 33,8% for Portuguese time series, respectively	Data collected from January 22th of 2020, until January 21th of 2021	Relative error	X
Motavali [41]	Introduction of the DSA-BEATS model, a novel hybrid forecasting approach	The model accurately predicts COVID-19 hospitalizations	Daily hospitalized COVID-19 cases, from 2020/08/04 to 2022/02/06	RMSE, MAPE	X
Li [45]	Development of a Transformer-based model, to increase the prediction capacity	The approach provides approximate performance in short-term forecasting and better performance in long-term forecasting	Patient counts from 47 prefectures in Japan, ranging from August 2012 to March 2020	RMSE, MAE, PC	Python
Moloshnikov [51]	Improve the accuracy of forecasting of COVID-19 in Russia for two years, using LSTM-Fully Connected	The considered models, both simple and more complex, have similar efficiency. The lowest error achieved is 18% MAPE for Moscow and 8% MAPE for Russia	Daily data on number of cases, recoveries and deaths from March 2021 to February 2022 for 85 regions of Russia and all Russia	MAPE	Python

Table 6
Mixed or Combined models.

1st Author	Main Idea/Objective	Main findings	Characteristics of dataset	Metrics	Language
Rasjid [25]	Predict the death and infected COVID-19 in Indonesia using Savitzky Golay Smoothing and LSTM Neural Network	LSTM-NN prediction produce better result than the Savitzky Golay Smoothing	147 records from 2 March 2020 to 26 July 2020. 121 data used for training, and the rest for testing	MSE, R^2	Python
Lobato [33]	Use of SVR, LSTM, and ANN to predict daily incidence and prevalence for nine countries in Latin America	There is no single best model or best hyperparameters configuration for all countries and targets	Data from 03/04/2020 to 07/03/2020, for countries in Latin America with a population of more than 10 million	Grid search, MAE	Python
Assad [36]	Apply 6 univariate and 2 multivariate models to evaluate 14 time series from a Brazilian city (RJ), all Italian regions and 4 US states	After comparison of univariate and multivariate methods, univariate models presented best results in almost all regions	Time-series data range of 369 days (March 4, 2020, to July 3, 2021), for Brazil, Italy and US	RMSE	R
Balli [40]	Time series prediction model is proposed to obtain the disease curve and forecast the epidemic trend using machine learning methods	SVM achieved the best trend. According to estimates, the global pandemic will peak at the end of January 2021	Data of weekly confirmed cases and cumulative confirmed cases for 35 weeks (20/01/2020 and 18/09/2020). 18 weeks for training and 17 weeks as test	RMSE, APE, MAPE	X
Liang [42]	Investigate the effectiveness of using web search queries in predicting the incidence of tuberculosis in China	BMA with weighted sample results achieved significantly lower RMSE and MAPE in all cities, provinces and the national aggregate	Monthly data of tuberculosis in 31 cities and provinces in China during 2004 to 2016	MAPE, RMSE	X
Swaraj [43]	Proposition of an ensemble model integrating ARIMA model to extract the linear correlations and a NAR-NN for modeling the residuals	The combination displayed significant reduction in RMSE (16.23%), MAE (37.89%) and MAPE (39.53%) values, when compared with single ARIMA model	Daily cumulative count of confirmed, deaths and recovered cases of COVID-19, between May 6–15, July 21–30 and Aug 1–10 2020	RMSE, MAE, MAPE	Matlab
Lucic [44]	Use of k-means to group US counties based on demographic and economic similarities; Then, use of time series forecasting of each cluster to assess the short-run viral transmissibility risk	Including sub-national socio-economic characteristics to data-driven COVID-19 forecasts play a key role in assessing the risk associated with changes in infection patterns at the national level	Demographic data for each county in the US; Daily cumulative COVID-19 cases and deaths from March to October 2020; Data on Gross Domestic Product (GDP) from 2010 to 2018; Data on GIS coordinates of each US county	RMSE, MAE	Python
Kumar [46]	Application of various machine learning, time series, and deep learning models to forecast COVID-19 cases	Random Forest Regressor, Facebook Prophet, and Stacked LSTM outperformed to predict best results for COVID-19 cases	The confirmed, deceased, and recovered datasets from January 22, 2020, to May 29, 2021, of Novel COVID-19 cases	RMSE, Rr	Python
Jin [48]	Combining three forecasting models to leverage the strengths of deep learning in capturing linear and nonlinear factors	The CNN-LSTM-ARIMA model exhibits the highest predictive accuracy, with an MSE of 7048.26, RMSE of 83.95, MAE of 61.18, MAPE of 0.16, and R^2 of 0.95	Daily COVID-19 cases from March 1, 2020 to Oct. 21, 2021 in Quebec, and the daily COVID-19 cases from Feb. 21, 2020 to Oct. 12, 2021 in Italy	MSE, MAE, RMSE, MAPE, R^2	Python
Chen [49]	Three countries data with more severe COVID-19 epidemics, to predict new cases in the next 15 days with LSTM, and use of Bayesian Optimization	Recursive Prediction LSTM is the best model, followed by Encoder-Decoder LSTM. Bayesian Optimizer help finding good hyperparameters	Official statistics from India, Russia and Chile, and number of new cases from the outbreak of the COVID-19 to Dec. 30, 2020	RMSE, MAPE	Python
Ayris [50]	Proposition of a deep learning technique (DSPM) and machine learning (NRM) to predict the spread of COVID-19	DSPM and NRM achieve best performance, respectively compared to TSM, LSTM-based DDEM and MHEM	Confirmed, recovery and death cases on 6.4 million COVID-19 cases, reported from 22nd January to 6th June 2020	MAE	X

(continued on next page)

Table 6 (continued).

Jin [53]	Combination of three models using data from the COVID-19 outbreak that affected Germany and Japan	The backpropagation neural network coupled to LSTM model and ARIMA model offers the highest accurate prediction effect	Total of infected data from Japan and Germany from April 1, 2020 to March 9, 2023	MSE, RMSE, MAE	X
Zrieq [58]	Analyze the historical of COVID-19 in order to construct robust and accurate forecasting models	Both models were found to be accurate and robust in forecasting the time series of COVID-19 in Saudi Arabia	Daily number of confirmed, recovered and death cases from 2 March 2020 to 22 June 2022	MAE, RMSE, R^2	Python
Punyaporn-withaya [59]	Model and forecast the monthly number of FMD outbreak episodes (n-FMD episodes) in Thailand using multiple time-series methods	n-FMD episodes had a stable trend from 2010 to 2020, and increased from 2014 to 2020, the outbreak episodes occurring from September to November annually	1209 FMD cases from Jan. 2010 to Dec. 2020, divided into two parts: from 2010 to 2019 for training and from 2020 for validation	RMSE, MAE, MASE	R
Olukanmi [60]	Models based on deep learning, machine learning, and Time series methods, using GTD to forecast ILI in South Africa	Deep learning techniques perform better, and Combining GT and historical ILI data enhances the models	Weekly date from 1st week of 2010 to the 43rd week of 2018 (459 weeks)	RMSE, MAE, PCC, PWD, PMD	Python

Table 7
Method category.

Mixed	Deep Learning	Machine Learning	Time series models
15	12	6	3

Table 8
Mixed methods.

TSM & DL	TSM & ML	TSM & DL ML	DL & ML
5	3	3	4

Table 9
Most used time series models (at least twice).

Es	LSTM	ARIMA	MLP	ANN	RNN	LR	SVR/ML
5	14	9	3	2	2	3	7
CNN	RF	SARIMA	Bayesian	NAR	KNN	FP	TBATS
2	3	5	4	2	2	2	2

Table 10
Evaluation metrics.

PPV	NVP	F1	MAPE	RMSE	MAD	R^2	MSE
1	1	2	14	24	1	6	5
AIC	BIC	MAE	AUC	PCC	PMD	PWD	
1	2	14	1	2	1	1	

Table 11
Language/Tool.

Python	R	Matlab	Weka	AUtoML	Not specified
15	4	2	1	1	13

solely relied on time series models, predominantly ARIMA/SARIMA methods. However, starting from 2020, the application of TSMS surged, especially with the onset of the COVID-19 pandemic, accompanied by a significant adoption of Machine Learning (ML) and Deep Learning (DL) techniques for epidemic prediction. This trend towards amalgamating various methodologies aims to capitalize on the strengths of each, ultimately enhancing forecasting accuracy. Despite the effectiveness of employing diverse time series techniques, there exists a pressing need for improved data collection, structuring and storage, particularly

highlighted by the challenges posed during the COVID-19 pandemic. This necessity has spurred a proliferation of scientific studies in this domain.

In light of these findings, we advocate for harnessing the potential of time series techniques to monitor and forecast epidemics globally, including emerging cases. To achieve this, leveraging diverse databases housing up-to-date datasets for various diseases and regions, such as CDC, JHUC, Google, Kaggle, and WHO, is imperative. Furthermore, programming languages like Python, equipped with extensive libraries,

Table 12

Main data source.

JHUC	CDC	Google	WHO
6	7	4	4

Table 13

Journal.

Science Direct	IEEE	MDPI	ACM	Springer	Elsevier	PLoS ONE	Wiley
5	9	3	8	2	6	2	1

Table 14

Publication per year.

2023	2022	2021	2020	2019	2018	2014
6	8	12	6	2	1	1

Table 15

Disease addressed.

COVID-19	Influenza	Diarrhea	Dengue fever	Tuberculosis	HFMD
26	4	1	2	1	2

Table 16

Country (at least 3 works on the Country).

Colombia	Canada	Germany	China	India	Italy	Russia	Brazil	USA
3	3	4	4	6	7	7	9	13

Table 17

Continent.

Africa	Asia	America	Europe
5	3	3	4

offer immense potential for constructing robust models and integrating them to bolster epidemic surveillance and prediction efforts.

The unique epidemiological landscape of Africa warrants special attention due to the frequent occurrence of epidemics.

The COVID-19 pandemic has underscored the efficacy of time series-based methods in addressing both public health and clinical cases, highlighting the imperative to capitalize on this momentum to further research and development efforts for combating prevalent diseases.

Consequently, this SLR reaffirms our commitment to using time series to contribute to epidemic management and clinical health cases. As part of ongoing initiatives, we have embarked on several projects leveraging data from the Health Information Unit of the Cameroon Ministry of Public Health [62,63]. In future research endeavours, we propose incorporating network models, specifically drawing inspiration from seminal works such as [11,64–67].

While our SLR provides valuable insights, it is not without limitations. The number and quality of databases used could be enhanced, and the study period extended to examine the impact of other epidemics on research. Additionally, future research could expand to encompass analysis methods beyond forecasting techniques.

5. Conclusion

This study aimed to assess the efficacy of Time Series Methods (TSMs) in predicting epidemics and anticipated case numbers. Employing the SPIDER tool to define our research question, delineating specific objectives, and following the PRISMA methodology, we surveyed four scientific databases. Through a rigorous application of inclusion and exclusion criteria and meticulous screening, we initially identified 3325 works, ultimately selecting 36 pertinent studies for inclusion in our Systematic Literature Review (SLR). Our analysis unveiled that 2021 witnessed the highest volume of publications, totaling 12, with Deep Learning (DL) methods emerging as the most prevalent, featuring in 12 papers. However, there is a discernible shift towards integrating various

methodologies, evident in 15 papers. Notably, the combination of Time Series Models and DL proved popular, featured in 5 works. Leading journals such as IEEE and ACM showcased a considerable presence, hosting 9 and 8 articles, respectively. Predominant data sources included CDC, JHUC, Google, and WHO. Python emerged as the primary programming language, utilized in 15 papers. Remarkably, the United States and Brazil emerged as the focal points for research endeavors, hosting 13 and 9 studies, respectively, positioning the Americas as the continent with the highest research activity. Among the plethora of time series methods employed, Long Short-Term Memory (LSTM) and ARIMA/SARIMA emerged as the most prevalent, each featured in 14 works. Evaluation metrics predominantly comprised RMSE, followed by MAPE and MEA. Intriguingly, only a fraction of works addressed missing and outlier data management, indicating areas for potential improvement. Notably, few studies have addressed the African context, despite the region's vulnerability to epidemics. The advent of the COVID-19 pandemic in late 2019 catalyzed research in time series applications in epidemiology, with 26 out of 36 studies focusing on COVID-19. This SLR underscores the imperative to advance research endeavors, particularly in the African context, emphasizing the need to incorporate pertinent data and diseases endemic to the region.

CRedit authorship contribution statement

Apollinaire Batoure Bamana: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Mahdi Shafiee Kamalabad:** Writing – review & editing, Writing – original draft, Methodology. **Daniel L. Oberski:** Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table 18
Cross-analysis of methods by year of use.

	2023	2022	2021	2020	2019	2018	2014	Total
Mixed	03	04	05	03	X	X	X	15
DL	02	01	07	01	01	X	X	12
ML	01	03	X	02	X	X	X	06
TSM	X	X	X	X	01	01	01	03
Total	06	08	12	06	02	01	01	36

Table 19
Work done on COVID-19, by year.

	2023	2022	2021	2020	Total
COVID-19	06	07	08	05	26

Acknowledgments

We thank the Staff of the Health Information Unit of the Cameroon Ministry of Public Health, for giving us access to their data base. We also thank The University of Ngooundere (Cameroun) for allowing us to have a scientific mobility. We finally thank Utrecht University (The Netherlands) who warmly welcomes us during the research stay.

Funding

This research was conducted during a scientific trip at Utrecht University, funded by the Coimbra Group through his program of Short Stay for young researchers from Sub-Saharan Africa, for the year 2022/2023.

References

- Petropoulos F. Forecasting: Theory and practice. *Int J Forecast* 2022;38(3):705–871. <http://dx.doi.org/10.1016/j.ijforecast.2021.11.001>.
- Makridakis S, Assimakopoulos V, Spiliotis E. Objectivity, reproducibility and replicability in forecasting research. *Int J Forecast* 2018;34(4):835–8. <http://dx.doi.org/10.1016/j.ijforecast.2018.05.001>.
- Brockwell PJ, Davis RA. Introduction to time series and forecasting. Springer texts in statistics, Cham: Springer International Publishing; 2016. <http://dx.doi.org/10.1007/978-3-319-29854-2>.
- Caporin M, Storti G. Financial time series: Methods and models. *J Risk Financ Manag* 2020;13(5). <http://dx.doi.org/10.3390/jrfm13050086>.
- Mills TC. *Applied time series analysis: A practical guide to modeling and forecasting*. London: Academic Press; 2019.
- Hyndman RJ, Athanasopoulos G. *Forecasting: Principles and practice. Third print edition*. Melbourne, Australia: Otexts, Online Open-Access Textbooks; 2021.
- Smyl S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *Int J Forecast* 2020;36(1):75–85. <http://dx.doi.org/10.1016/j.ijforecast.2019.03.017>.
- Paoletta MS. *Linear models and time-series analysis: Regression, ANOVA, ARMA and GARCH*. Hoboken: John Wiley & Sons; 2019.
- Tealab A. Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Comput Inform J* 2018;3(2):334–40. <http://dx.doi.org/10.1016/j.fcij.2018.10.003>.
- Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep learning for time series forecasting: A survey. *Big Data* 2021;9(1):3–21. <http://dx.doi.org/10.1089/big.2020.0159>.
- Shafiee Kamalabad M, Grzegorzczak M. A new Bayesian piecewise linear regression model for dynamic network reconstruction. *BMC Bioinform* 2021;22(S2):196. <http://dx.doi.org/10.1186/s12859-021-03998-9>.
- Zhang Q. Data science approaches to infectious disease surveillance. *Philos Trans. Ser A, Math, Phys, Eng Sci* 2022;380(2214):20210115. <http://dx.doi.org/10.1098/rsta.2021.0115>.
- Brownlee J. *Deep learning for time series forecasting: Predict the future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery*; 2018.
- Wang Y, Yan Z, Wang D, Yang M, Li Z, Gong X, et al. Prediction and analysis of COVID-19 daily new cases and cumulative cases: Times series forecasting and machine learning models. *BMC Infect Dis* 2022;22(1):495. <http://dx.doi.org/10.1186/s12879-022-07472-6>.
- Guo Z, Zhang L, Liu J, Liu M. Impact of COVID-19 prevention and control on the influenza epidemic in China: A time series study. *Health Data Sci* 2022;2022:9830159. <http://dx.doi.org/10.34133/2022/9830159>.
- Hswen Y, Brownstein JS, Xu X, Yom-Tov E. Early detection of COVID-19 in China and the USA: summary of the implementation of a digital decision-support and disease surveillance tool. *BMJ Open* 2020;10(12). <http://dx.doi.org/10.1136/bmjopen-2020-041004>.
- Poletto C, Boëlle P-Y, Colizza V. Risk of MERS importation and onward transmission: A systematic review and analysis of cases reported to WHO. *BMC Infect Dis* 2016;16(1):448. <http://dx.doi.org/10.1186/s12879-016-1787-5>.
- Dogan O, Tiwari S, Jabbar MA, Guggari S. A systematic review on AI/ML approaches against COVID-19 outbreak. *Complex Intell Syst* 2021;7(5):2655–78. <http://dx.doi.org/10.1007/s40747-021-00424-8>.
- Keshavamurthy R, Dixon S, Pazdernik KT, Charles LE. Predicting infectious disease for biopreparedness and response: A systematic review of machine learning and deep learning approaches. *One Health* 2022;15:100439. <http://dx.doi.org/10.1016/j.onehlt.2022.100439>.
- Santangelo OE, Gentile V, Pizzo S, Giordano D, Cedrone F. Machine learning and prediction of infectious diseases: A systematic review. *Mach Learn Knowl Extraction* 2023;5(1):175–98. <http://dx.doi.org/10.3390/make5010013>.
- Leung XY, Islam RM, Adhami M, Ilic D, McDonald L, Palawaththa S, et al. A systematic review of dengue outbreak prediction models: Current scenario and future directions. *PLOS Negl Trop Dis* 2023;17(2):e0010631. <http://dx.doi.org/10.1371/journal.pntd.0010631>.
- Ajagbe SA, Adigun MO. Deep learning techniques for detection and prediction of pandemic diseases: A systematic literature review. *Multimedia Tools Appl* 2024;83(2):5893–927. <http://dx.doi.org/10.1007/s11042-023-15805-z>.
- Cooke A, Smith D, Booth A. Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qual Health Res* 2012;22(10):1435–43. <http://dx.doi.org/10.1177/1049732312452938>.
- Puckett J. *Zotero: A guide for librarians, researchers, and educators. 2nd ed.* Chicago, Illinois: Association of College and Research Libraries, a division of the American Library Association; 2017.
- Rasjid ZE, Setiawan R, Effendi A. A comparison: Prediction of death and infected covid-19 cases in Indonesia using time series smoothing and LSTM neural network. *Procedia Comput Sci* 2021;179:982–8. <http://dx.doi.org/10.1016/j.procs.2021.01.102>.
- Borghini PH, Zakordonets O, Teixeira JP. A COVID-19 time series forecasting model based on MLP ANN. *Procedia Comput Sci* 2021;181:940–7. <http://dx.doi.org/10.1016/j.procs.2021.01.250>.
- Venna SR, Tavanaei A, Gottumukkala RN, Raghavan VV, Maida AS, Nichols S. A novel data-driven model for real-time influenza forecasting. *IEEE Access* 2019;7:7691–701. <http://dx.doi.org/10.1109/ACCESS.2018.2888585>.
- Safari A, Hosseini R, Mazinani M. A novel deep interval type-2 fuzzy LSTM (DIT2FLSTM) model applied to COVID-19 pandemic time-series prediction. *J Biomed Inform* 2021;123:103920. <http://dx.doi.org/10.1016/j.jbi.2021.103920>.
- Yang E, Park HW, Choi YH, Kim J, Munkhdalai L, Musa I, et al. A simulation-based study on the comparison of statistical and time series forecasting methods for early detection of infectious disease outbreaks. *Int J Environ Res Public Health* 2018;15(5). <http://dx.doi.org/10.3390/ijerph15050966>. Place: Switzerland.
- Ferdousi T, Cohnstaedt LW, Scoglio CM. A windowed correlation-based feature selection method to improve time series prediction of dengue fever cases. *IEEE Access* 2021;9:141210–22. <http://dx.doi.org/10.1109/ACCESS.2021.3120309>.
- Nuanhuay T, Sinthupinyo S. Additional time series features for preciseness improvement of LSTM-based COVID-19 spread forecasting model. In: proceedings of the 2021 4th international conference on machine learning and machine intelligence. MLMI '21, New York, NY, USA: Association for Computing Machinery; 2022, p. 145–50. <http://dx.doi.org/10.1145/3490725.3490747>, event-place: Hangzhou, China.
- Liapis CM, Karanikola A, Kotsiantis S. An ensemble forecasting method using univariate time series COVID-19 data. In: Proceedings of the 24th pan-hellenic conference on informatics. New York, NY, USA: Association for Computing Machinery; 2021, p. 50–2. <http://dx.doi.org/10.1145/3437120.3437273>, event-place: Athens, Greece.

- [33] Lobato Barreira B, Fray da Silva R, Eduardo Cugnasca C. An in-depth analysis on the use of long short-term memory networks to predict incidence and prevalence of covid-19 in Latin America. In: 2021 international symposium on electrical, electronics and information engineering. New York, NY, USA: Association for Computing Machinery; 2021, p. 384–90. <http://dx.doi.org/10.1145/3459104.3459167>, event-place: Seoul, Republic of Korea.
- [34] Watmaha J, Kamonsantiroj S, Pipanmaekaporn L. An integrated climate and spatio-temporal determinant for influenza forecasting based on convolution neural network. In: Proceedings of the 9th international conference on computer and communications management. ICCCM '21, New York, NY, USA: Association for Computing Machinery; 2021, p. 104–10. <http://dx.doi.org/10.1145/3479162.3479178>, event-place: Singapore, Singapore.
- [35] Tetteroo J, Baratchi M, Hoos HH. Automated machine learning for COVID-19 forecasting. *IEEE Access* 2022;10:94718–37. <http://dx.doi.org/10.1109/ACCESS.2022.3202220>.
- [36] Assad DBN, Cara J, Ortega-Mier M. Comparing short-term univariate and multivariate time-series forecasting models in infectious disease outbreak. *Bull Math Biol* 2022;85(1):9. <http://dx.doi.org/10.1007/s11538-022-01112-5>, Place: United States.
- [37] Masum AKM, Khushbu SA, Keya M, Abujar S, Hossain SA. COVID-19 in Bangladesh: A deeper outlook into the forecast with prediction of upcoming per day cases using time series. *Procedia Comput Sci* 2020;178:291–300. <http://dx.doi.org/10.1016/j.procs.2020.11.031>.
- [38] Kim Y, Park C-R, Ahn J-P, Jang B. COVID-19 outbreak prediction using Seq2Seq + Attention and Word2Vec keyword time series data. *PLoS One* 2023;18(4):e0284298. <http://dx.doi.org/10.1371/journal.pone.0284298>, Place: United States.
- [39] Carvalho KCMd, Vicente JP, Teixeira JP. COVID-19 time series forecasting – twenty days ahead. *Procedia Comput Sci* 2022;196:1021–7. <http://dx.doi.org/10.1016/j.procs.2021.12.105>.
- [40] Balli S. Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. *Chaos, Solitons, Fractals* 2021;142:110512. <http://dx.doi.org/10.1016/j.chaos.2020.110512>, Place: England.
- [41] Motavali A, Yow K-C, Hansmeier N, Chao T-C. DSA-BEATS: Dual self-attention N-BEATS model for forecasting COVID-19 hospitalization. *IEEE Access* 2023;11:137352–65. <http://dx.doi.org/10.1109/ACCESS.2023.3318931>.
- [42] Liang X, Xu Q, Guan R, Zhao Y. Forecasting tuberculosis incidence in China using Baidu index: A comparative study. In: Proceedings of the 4th international conference on medical and health informatics. New York, NY, USA: Association for Computing Machinery; 2020, p. 22–9. <http://dx.doi.org/10.1145/3418094.3418129>, event-place: Kamakura City, Japan.
- [43] Swaraj A, Verma K, Kaur A, Singh G, Kumar A, Sales LMD. Implementation of stacking based ARIMA model for prediction of Covid-19 cases in India. *J Biomed Inform* 2021;121:103887. <http://dx.doi.org/10.1016/j.jbi.2021.103887>.
- [44] Lucic MC, Ghazzai H, Lipizzi C, Massoud Y. Integrating county-level socioeconomic data for COVID-19 forecasting in the united states. *IEEE Open J Eng Med Biol* 2021;2:235–48. <http://dx.doi.org/10.1109/OJEMB.2021.3096135>.
- [45] Li L, Jiang Y, Huang B. Long-term prediction for temporal propagation of seasonal influenza using Transformer-based model. *J Biomed Inform* 2021;122:103894. <http://dx.doi.org/10.1016/j.jbi.2021.103894>.
- [46] Kumar Y, Koul A, Kaur S, Hu Y-C. Machine learning and deep learning based time series prediction and forecasting of ten nations' COVID-19 pandemic. *SN Comput Sci* 2023;4(1):91. <http://dx.doi.org/10.1007/s42979-022-01493-3>, Place: Singapore.
- [47] Irungu J, Oladunni T, Grizzle AC, Denis M, Savadkoochi M, Ososanya E. ML-EKG-COVID: A machine learning-electrocardiogram signal processing technique for COVID-19 predictive modeling. *IEEE Access* 2023;11:135994–6014. <http://dx.doi.org/10.1109/ACCESS.2023.3335384>.
- [48] Jin Y-C, Cao Q, Sun Q, Lin Y, Liu D-M, Shan-Yu, et al. Models for COVID-19 data prediction based on improved LSTM-ARIMA algorithms. *IEEE Access* 2024;12:3981–91. <http://dx.doi.org/10.1109/ACCESS.2023.3347403>.
- [49] Chen T, Chen S, Mei S, An S, Yuan X, Lu Y. Multistep forecasting of new COVID-19 cases based on LSTMs using Bayesian optimization. In: 2021 international symposium on electrical, electronics and information engineering. New York, NY, USA: Association for Computing Machinery; 2021, p. 55–61. <http://dx.doi.org/10.1145/3459104.3459116>, event-place: Seoul, Republic of Korea.
- [50] Ayris D, Imtiaz M, Horbury K, Williams B, Blackney M, See CSH, et al. Novel deep learning approach to model and predict the spread of COVID-19. *Intell Syst Appl* 2022;14:200068. <http://dx.doi.org/10.1016/j.iswa.2022.200068>.
- [51] Moloshnikov IA, Shboev AG, Naumov AV, Zaverlyayev SV, Rybka RB. On the accuracy of Covid-19 forecasting methods in Russia for two years. *Procedia Comput Sci* 2022;213:428–34. <http://dx.doi.org/10.1016/j.procs.2022.11.088>.
- [52] Jain P, Sahu S. Prediction and forecasting of worldwide corona virus (COVID-19) outbreak using time series and machine learning. *Concurr Comput : Pract Exp* 2022;e7286. <http://dx.doi.org/10.1002/cpe.7286>, Place: England.
- [53] Jin Y-C, Cao Q, Wang K-N, Zhou Y, Cao Y-P, Wang X-Y. Prediction of COVID-19 data using improved ARIMA-LSTM hybrid forecast models. *IEEE Access* 2023;11:67956–67. <http://dx.doi.org/10.1109/ACCESS.2023.3291999>.
- [54] Susarla A, Liu A, Thai DH, Le MT, Züfle A. Spatiotemporal disease case prediction using contrastive predictive coding. In: Proceedings of the 3rd ACM SIGSPATIAL international workshop on spatial computing for epidemiology. New York, NY, USA: Association for Computing Machinery; 2022, p. 26–34. <http://dx.doi.org/10.1145/3557995.3566122>, event-place: Seattle, Washington.
- [55] Salgotra R, Gandomi M, Gandomi AH. Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. *Chaos, Solitons Fractals* 2020;138:109945. <http://dx.doi.org/10.1016/j.chaos.2020.109945>, Place: England.
- [56] Somboonsak P. Time series analysis of dengue fever cases in thailand utilizing the SARIMA model. In: Proceedings of the 2019 7th international conference on information technology: IoT and smart city. New York, NY, USA: Association for Computing Machinery; 2020, p. 439–44. <http://dx.doi.org/10.1145/3377170.3377215>, event-place: Shanghai, China.
- [57] Feng H, Duan G, Zhang R, Zhang W. Time series analysis of hand-foot-mouth disease hospitalization in Zhengzhou: Establishment of forecasting models using climate variables as predictors. *PLoS One* 2014;9(1):e87916. <http://dx.doi.org/10.1371/journal.pone.0087916>, Place: United States.
- [58] Zriek R, Kamel S, Boubaker S, Algahtani FD, Alzain MA, Alshammari F, et al. Time-series analysis and healthcare implications of COVID-19 pandemic in Saudi Arabia. *Healthcare (Basel, Switzerland)* 2022;10(10). <http://dx.doi.org/10.3390/healthcare10101874>, Place: Switzerland.
- [59] Punyapornwithaya V, Mishra P, Sansamur C, Pfeiffer D, Arjkumpa O, Prakotcheo R, et al. Time-series analysis for the number of foot and mouth disease outbreak episodes in cattle farms in thailand using data from 2010–2020. *Viruses* 2022;14(7). <http://dx.doi.org/10.3390/v14071367>, Place: Switzerland.
- [60] Olukanmi SO, Nelwamondo FV, Nwulu NI. Utilizing Google search data with deep learning, machine learning and time series modeling to forecast influenza-like illnesses in South Africa. *IEEE Access* 2021;9:126822–36. <http://dx.doi.org/10.1109/ACCESS.2021.3110972>.
- [61] Kong W-H, Li Y, Peng M-W, Kong D-G, Yang X-B, Wang L, et al. SARS-CoV-2 detection in patients with influenza-like illness. *Nature Microbiol* 2020;5(5):675–8. <http://dx.doi.org/10.1038/s41564-020-0713-1>.
- [62] Moskalai Ngossaha J, Ynsufu A, Batoure Bamana A, Djeumen R, Bowong Tsakou S, Ayissi Eteme A. Towards a flexible urbanization based approach for integration and interoperability in heterogeneous health information systems: case of cameroon. In: Tchakounte F, Atemkeng M, Rajagopalan R, editors. Safe, Secure, Ethical, Responsible Technologies and Emerging Applications. SAFER-TEA 2023. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 566. Cham: Springer; 2024. http://dx.doi.org/10.1007/978-3-031-56396-6_16.
- [63] Batoure Bamana A, Dangbe E, Abboubakar H, Shafiee Kamalabad M. A comprehensive statistical analysis of malaria dynamics in the Adamawa Region of Cameroon, from 2018 to 2022. *Brazilian Journal of Biometrics* 2024;42(3). <http://dx.doi.org/10.28951/bjb.v42i3.703>.
- [64] Shafiee Kamalabad M, Grzegorzczak M. Improving nonhomogeneous dynamic Bayesian networks with sequentially coupled parameters. *Stat Neerl* 2018;72(3):281–305. <http://dx.doi.org/10.1111/stan.12136>.
- [65] Shafiee Kamalabad M, Heberle AM, Thedieck K, Grzegorzczak M. Partially non-homogeneous dynamic Bayesian networks based on Bayesian regression models with partitioned design matrices. *Bioinformatics* 2019;35(12):2108–17. <http://dx.doi.org/10.1093/bioinformatics/btz917>.
- [66] Shafiee Kamalabad M, Grzegorzczak M. Non-homogeneous dynamic Bayesian networks with edge-wise sequentially coupled parameters. *Bioinformatics* 2020;36(4):1198–207. <http://dx.doi.org/10.1093/bioinformatics/btz690>.
- [67] Shafiee Kamalabad M, Leenders R, Mulder J. What is the point of change? Change point detection in relational event models. *Social Networks* 2023;74:166–81. <http://dx.doi.org/10.1016/j.socnet.2023.03.004>.