REVIEW ARTICLE

Tracking machine learning models for pandemic scenarios: a systematic review of machine learning models that predict local and global evolution of pandemics

Marcelo Benedeti Palermo[1] · Lucas Micol Policarpo[1] · Cristiano André da Costa[1] · Rodrigo da Rosa Righi[1]

Abstract

This systematic review aims to study and classify machine learning models that predict pandemics' evolution within affected regions or countries. The advantage of this systematic review is that it allows the health authorities to decide what prediction model fits best depending upon the region's criticality and optimize hospitals' approaches to preparing and anticipating patient care. We searched ACM Digital Library, Biomed Central, BioRxiv+MedRxiv, BMJ, Computers and Applied Sciences, IEEEXplore, JMIR Medical Informatics, Medline Daily Updates, Nature, Oxford Academic, PubMed, Sage Online, ScienceDirect, Scopus, SpringerLink, Web of Science, and Wiley Online Library between 1 January 2020 and 31 July 2022. We divided the interventions into similarities between cumulative COVID-19 real cases and machine learning prediction models' ability to track pandemics trending. We included 45 studies that rated low to high risk of bias. The standardized mean differences (SMD) for the two groups were 0.18, 95% CI, with interval of [0.01, 0.35], $I^2=0$, and p value=0.04. We built a taxonomic analysis of the included studies and determined two domains: pandemics trending prediction models and geolocation tracking models. We performed the meta-analysis and data synthesis and got low publication bias because of missing results. The level of certainty varied from very low to high. By submitting the 45 studies on the risk of bias, the levels of certainty, the summary of findings, and the statistical analysis via the forest and funnel plots assessments, we could determine the satisfactory statistical significance homogeneity across the included studies to simulate the progress of the pandemics and help the healthcare authorities to take preventive decisions.

Keywords Machine learning models · Pandemic prediction · Systematic review · Meta-analysis

Marcelo Benedeti Palermo, Lucas Micol Policarpo, and Cristiano André da Costa have contributed equally to this work.

✉ Cristiano André da Costa
cac@unisinos.br

Marcelo Benedeti Palermo
mbpalermo@edu.unisinos.br

Lucas Micol Policarpo
lmicolpolicarpo@unisinos.br

Rodrigo da Rosa Righi
rrrighi@unisinos.br

[1] Software Innovation Laboratory-SOFTWARELAB, Programa de Pós-Graduação em Computação Aplicada, Universidade do Vale do Rio dos Sinos, Av. Unisinos 950, São Leopoldo, RS 93022-750, Brazil

1 Introduction

1.1 The social motivation

Pandemic scenarios are always challenging, because they need massive, fast, and practical actions to save lives. The SARS-CoV-2 pandemics have become a unique event in the twenty-first century of humankind, surpassing all pandemic effects caused by the Spanish flu in 1918. All social complexities and gaps in a world where the current population walks toward the 8 billion inhabitants mark overcome all realities of a 1.5 billion crowded world from 1918. The world's inhabitants from 1918 did not have effective vaccines or medications to fight against the H1N1 virus. Such a combination resulted in around 50 million deaths, and one-third of the world population was infected (Jordan 2019).

During COVID-19 pandemics, developed nations keep at least 50% of all global vaccine doses available and leave

emerging countries behind, increasing the chances of the virus spreading across a population (Fox 2020).

1.2 The biological motivation

The chances of novel virus variants surging (Gallagher 2021) increase significantly when the vaccination process covers only a tiny parcel of the population. The natural virus variance mechanism is always fast enough to find more efficient ways to hijack human host cells and incubate and replicate faster. The result can be catastrophic and cause the pandemic scenario to go out of control of healthcare authorities (Shang et al. 2020). Once in this stage, it may become hard to retake control of the health situation in a region affected by an outbreak. That is why, virus contact tracing plays a critical role in reducing infections. However, virus contact tracing is an arduous manual task and is not enough when a novel virus jumps into humans (Buhat et al. 2021; Shufro 2020).

Both social and biological aspects have motivated the scientific community to find fast responses to protect the population since 2020. Applied computing has been collaborating with the health area by providing computational models that can contribute by attempting to predict the next steps ahead of the pandemic.

Ghaderzadeh and Asadi (2021) provide a systematic review of the current state of all models for detecting and diagnosing COVID-19 through radiology modalities and their processing based on deep learning (DP). Alyasseri et al. (2022) provided a comprehensive review of the most recent DL and machine learning (ML) techniques for COVID-19 diagnosis. Syeda et al. (2021) conducted a systematic review of the literature on the role of artificial intelligence (AI) as a technology to fight the COVID-19 crisis in the fields of epidemiology, diagnosis, and disease progression.

1.3 Objective

Indeed, all previous systematic reviews cited in Sect. 1.2 are critical guides to helping the health area to thrive during the pandemic. However, our systematic review presents an essential advantage to help health authorities to save essential time and be steps ahead of the disease. This systematic review aims to review ML model studies that track pandemic scenario trends within an affected region compared to the actual trends in the same region. The goal is to answer the question: how effectively do ML models behave compared to the reality in the regions affected by the pandemics? The results obtained in this systematic review will work as a reference for healthcare authorities to make fast decisions and position steps ahead of the pandemics, mitigate deaths, and prevent overcrowded hospital admissions. The scientific contribution of this systematic review is to use the lessons learned from the ML models to apply them appropriately in

future pandemic scenarios and be preventively successful for population health safety.

The usage of ML models helps attenuate out-of-control pandemic scenarios once the prediction matches possible critical situations with a grade of certainty. To do that, we need to analyze the works published where authors proposed ML model predictions and how close the ML approached real scenarios in territories where COVID-19 caused damage.

To organize the results of this systematic review, we will use the *Participants, Intervention, Comparison and Outcomes* (PICO) approach (Moher et al. 2009; Liberati et al. 2009).

2 Materials and methods

We performed a systematic review of the articles that analyze COVID-19 pandemic scenarios and computational models used for intervention, prediction, and mitigation of the critical aspects of the virus spreading to optimize the conduction of the pandemic phase by sanitary authorities.

The systematic review uses Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA 2020 statement) to ensure the necessary transparency, protocol application, and guidelines (Moher et al. 2009; Liberati et al. 2009; Page et al. 2021).

This section discusses the research design and the steps used to achieve the goal of this systematic review. This paper addresses the eligibility criteria, the information sources, the research questions, the study selection, the data collection, and the article selection process.

2.1 Eligibility criteria

We have specified inclusion and exclusion criteria as synthesis to reach the necessary grouped studies.

The search takes place in July 2022 to identify meta-analyses focusing on computational models used for intervention, prediction, and mitigation of SARS-CoV-2 pandemic scenarios published between January 2020 and July 2022.

We have defined and refined the search string using the PICO approach (Booth et al. 2019). PICO covers PRISMA checklist topics, such as objectives, search questions, and eligibility criteria, by clustering participants, interventions, comparisons, and outcomes to facilitate our searches.

- Participants: We included randomized controlled trials of the COVID-19-positive population within specific territories. We considered individual mobility within geolocation as a factor of virus transmissibility.
- Intervention: we considered all interventions identified by the search where a pre-defined algorithm or plan aims

to predict the following steps within the pandemic scenario and how such methods can improve critical scenarios. We do not consider analyses and models that do not offer options to sanitary authorities.

- Comparison: We assess the studies where ML prediction models for pandemic scenarios can help sanitary authorities map and plan necessary steps to control, mitigate, or prevent the virus from spreading.
- Outcomes: To be included, a trial had to use ML or mathematical models (MM) that consider the COVID-19-positive population, the territory in scope, the mobility of individuals (whenever present), geolocation information (or a possibility to consider such), the power of ML or MM to focus on helping sanitary authorities management to tackle the scenario and to take over the control over the situation again (examples: prediction of future virus waves, preventive actions like lockdown or restrictions to slow down the virus spreading, or what vaccines are more effective depending upon the virus variants detected in a studied territory). We exclude reports with analyses of ML models without aiming to help sanitary authorities (examples: surveys about methods in ML, MM focused only on simulation, but without a goal of efficient actions toward population).

We considered articles written in English only. We performed the search and retrieval in all databases on July 2022. We excluded theses, dissertations, opinions, criticism, protocols, books, posters, oral presentations, general surveys, and previous non-related systematic and literature reviews.

We present the detailed inclusion and exclusion criteria established in this systematic review below:

- Inclusion criteria: articles that use ML to attempt to build or identify a critical pandemics region due to a significant virus spreading; articles considering geolocation studies and implementations within a pandemic territory study; articles using reinforcement learning as an alternative to help mitigate the pandemics; articles that are focusing on pandemics prediction or evolution within a specific territory using ML; articles considering machine or specific deep learning prediction approaches intended to help sanitary authorities; articles concerning SARS-CoV-2 variants and spreading trends over regions; articles considering population mobility between locations affected by SARS-CoV-2.
- Exclusion criteria: non-ML models, articles involving COVID-19 drug matching or finding; articles exploring ML closed to medical standpoints only, like medical imaging, X-ray, tomography to help identify COVID-19 presence; informative medical articles with superficial, generic, or overview medical interventions or analysis against SARS-CoV-2; articles exploring socioeconomic

Table 1 All databases covered in this systematic review

Database	Coverage
ACM Digital Library	January 2020 to July 2022
BMJ Journals	January 2020 to July 2022
BioRxiv + MedRxiv	January 2020 to July 2022
EBSCOhost	
Computers and Applied Sciences	January 2020 to July 2022
IEEE Xplore	January 2020 to July 2022
ElSevier	
ScienceDirect	January 2020 to July 2022
Scopus	January 2020 to July 2022
Journal of Medical Internet Research	
JMIR Medical Informatics	January 2020 to July 2022
Ovid	
MEDLINE(R) Daily Update	January 2020 to July 2022
Oxford Academic (all journals)	January 2020 to July 2022
NCBI PubMed PMC	January 2020 to July 2022
Sage Journals Online	January 2020 to July 2022
Springer Nature	
Biomed Central Journals	January 2020 to July 2022
Nature	January 2020 to July 2022
Springer Link	January 2020 to July 2022
Web of Science	January 2020 to July 2022
Wiley Online Library	January 2020 to July 2022

aspects, or blockchain methods during or after pandemics; reports about the usage of wearable IoT devices to help detect COVID-19 presence or exposure; articles focusing on population behavioral issues, mental health changes, or education impacts on the society during the pandemics; articles using algorithms to detect death tolls per se, without objectives to predict virus spreading or improvements in pandemic scenarios; articles about hospital capacity matching or optimization through ML algorithms (except when ML algorithms can also be used for epidemiological tracking); articles where bio environment aspects are involved as an effect of pandemics, like air pollution, climate change, or city water quality; general, direct, or indirect survey articles, which aim to cite or compare ML techniques as their focus.

2.2 Information sources

On 31 July 2022, we searched each of the databases listed in Table 1. We adapted the base search string from Table 2 to ensure that each database will return correct records to match our systematic review objectives (see Appendix A for details). We describe further information about the search strategy employed in this paper in Sect. 2.3. We queried the database by including the exact dates (01 January 2020 to 31 July 2022). However, some databases are

Table 2 Base search string we use during databases record retrieval

Search String
(COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning)

limited to the publication year only. For these cases, we included 2020 to 2022 and retrieved the available records until 31 July 2022.

2.3 Search strategy

To execute the search strategy, we need to define general and specific research questions to cover applications and results for the computational models and techniques found during the systematic review.

Here are the General Questions (GQ) addressed by this study:

- **GQ1.** What were the primary studies regarding ML COVID-19 models to track pandemics' evolution within a territory?
- **GQ2.** What are the challenges in algorithms and computational methods to help sanitary authorities control pandemic scenarios?

These are the Specific Questions (SQ) employed during this systematic review:

- **SQ1.** How do studies map COVID-19-positive population within the territory through ML models?
- **SQ2.** What studies consider SARS-CoV-2 virus variants tied to COVID-19-positive population within a specific territory?
- **SQ3.** What studies help healthcare authorities predict real-time or soft real-time pandemics evolution via ML models?
- **SQ4.** What studies focus on building geolocation maps based on the COVID-19-positive population?

We defined the following base search string based on general and specific questions to query all eligible and pertinent databases defined in Table 1.

We detail the individual search strings for each queried database in Appendix A.

The search strategy development process considered candidate search terms identified by looking at words in those records' titles, abstracts, and subject indexing. The databases strategy uses the Cochrane RCT filter reported in the Cochrane Handbook (Lasserson et al. 2019; Thomas et al. 2019).

2.4 Selection process

We combined two approaches to ensure accuracy during the selection process: *assessment of records by more than one reviewer* (Gartlehner et al. 2020; Waffenschmidt et al. 2019; Wang et al. 2020), and *crowdsourcing* (Noel-Storr 2019; Nama et al. 2019). In the case of crowdsourcing, we used the Mendeley Reference Manager (MRM) Tool as a central repository of records for teamwork. For this purpose, the tool has a specific feature named *Private Group*.

We did not use any ML approach to perform the selection process. Three researchers divided the screening tasks into abstracts, titles, keywords, and full-text articles. In case of inconsistencies or disagreements on such stages, the three researchers discussed until they had reached a consensus.

2.5 Data collection process

We have followed meticulous steps for the retrieval, selection, inclusion, and exclusion of records during this systematic review. To achieve the result, we executed the tasks detailed in the workflow from Fig. 1. We used the following tools to assist the data collection:

- Mendeley Reference Manager (MRM): the central repository of records collected through the search string from Table 2
- Mendeley Desktop Tool (MDT): the tool that removes duplicated records
- Mendeley Web Importer (MWB): the tool that imports records directly to MRM from the web search results pages for each database in Table 1

2.6 Data items

We organized the included studies in two domains and built a taxonomy to facilitate the analyses for the results section, according to Fig. 2.

All models and algorithms derived from ML techniques are eligible for inclusion. ML models focusing on predictions of pandemic spreading and geolocation will have preference during the inclusion process.

We collected data on:

- The report: author, year, and source of publication.
- The study: types of ML techniques used, sample applications

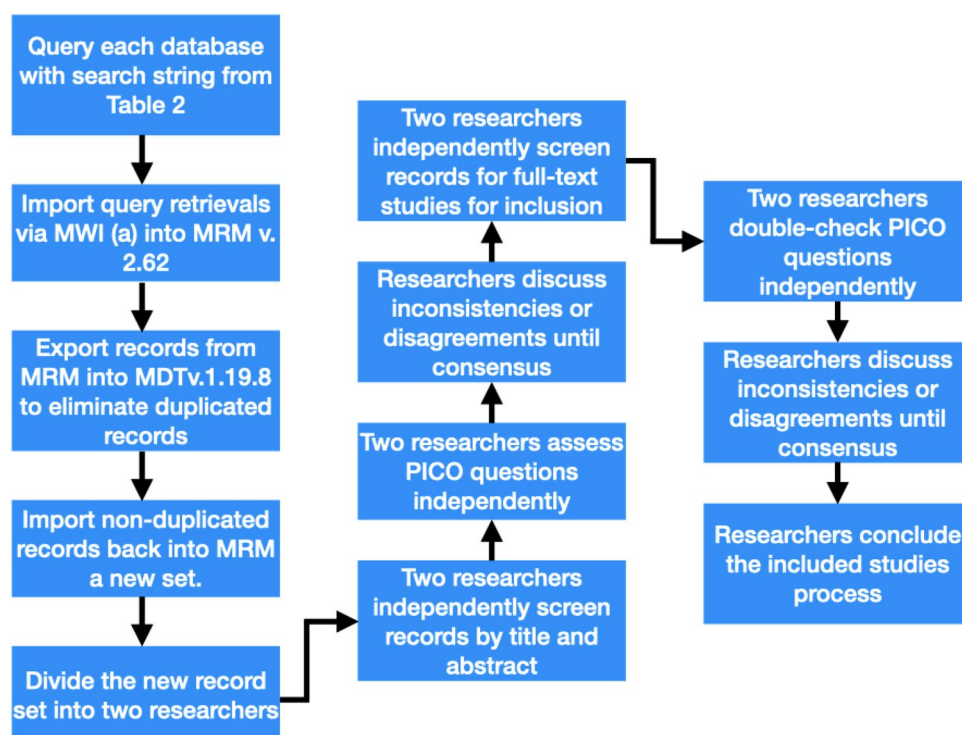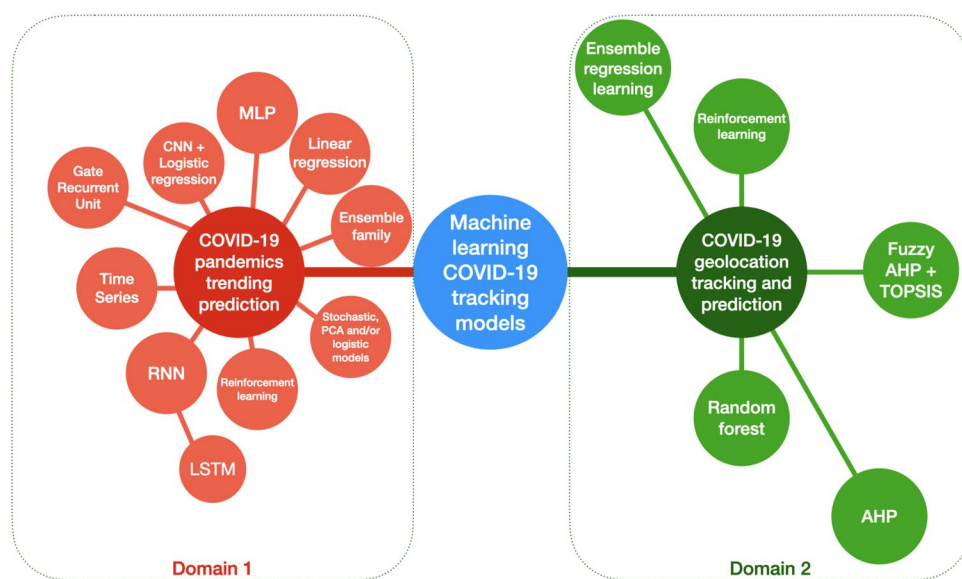**Fig. 1** Steps performed to execute the data collection process

Fig. 2 Taxonomic analysis and definition of (**a**) **Domain 1: COVID-19 pandemics' trending prediction** and (**b**) **Domain 2: COVID-19 geolocation tracking and prediction**

- The participants: COVID-19-positive individuals and their geographical location, whenever available
- The intervention: type, duration, timing, and mode of delivery of the employed ML model.

2.7 Risk of bias

We assessed the risk of bias (RoB) in the included studies by following the revised Cochrane risk-of-bias tool for randomized trials (RoB 2.0) (Higgins et al. 2019; Collaboration 2021). RoB 2.0 addresses five specific domains:

1. Bias arising from the randomization process
2. Bias due to deviations from intended intervention
3. Bias due to missing outcome data
4. Bias in measurement of the outcome
5. Bias in selection of the reported result.

Three review authors independently applied RoB 2.0 tool to each included study. They recorded supporting information and justifications for judgements of risk of bias for each domain (*low*, *high*, or *some concerns*).

We discussed and resolved all discrepancies in judgments of RoB or justifications for judgments to reach a consensus among the three review authors. By following the guidelines from RoB 2.0 tool (Sect. 1.3.4) (Higgins et al. 2019), we derived an overall summary risk of bias judgement (*low*, *high*, or *some concerns*) for each specific outcome.

The review authors determined RoB for each study through the highest RoB level in any of the assessed domains.

For all studies with missing outcomes, we proceeded with instructions from Sects. 10.12.2 and 10.12.3 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2019) and did not exclude the studies. Instead, we considered the effects of the missing outcome during the risk-of-bias determinations.

2.8 Effect measures

According to Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2019), our approach is to analyze the effect measures using the continuous data method when presenting the synthesis of the meta-analysis in this systematic review. We applied the effect size measures by first testing fixed-effect and random effects to determine the best statistical approach. We also adopted standardized means differences by summarizing the effect size measures within a forest plot.

Three review authors independently applied Cochrane Review Manager 5.4.1 tool (Collaboration 2020) to assess each included study and determine the effect size measure for all interventions and forest plot presentation. In case of any discrepancy or dual judgments, the three review authors resolved the conflict through discussion to reach a consensus.

2.9 Synthesis methods

To get the included studies that answer the researched questions, we categorized the research questions into the following main ML categories for the employed interventions:

1. COVID-19 pandemic trending prediction models
2. COVID-19 geolocation tracking and prediction models.

This approach helped us build a corresponding taxonomy map to facilitate the answers to the research questions framework and make the *Grading of Recommendations Assessment, Development, and Evaluation* (GRADE) summary of findings (Schunemann 2013).

Within the GRADE summary findings, we could assess the risk of bias and level of certainty within the 95% confidence interval of grouped and pooled results.

We used the synthesis methods described in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2019). For this systematic review, we used the following synthesis meta-analysis aspects regards effect measures:

1. We divided the meta-analysis into two comparison groups: actual COVID-19 cases' detection versus ML COVID-19 cases' prediction.
2. We got study heterogeneity $I^2 = 0$ in random-effects analysis. Section 10.10.4.1 of Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2019) states we should never choose between a fixed-effects and a random-effects meta-analysis based on the statistical test for heterogeneity only. However, we tested both fixed-effects and random-effects meta-analysis, and both approaches returned the same pooled results within the 95% confidence interval. Such developments and statistical significance led us to choose a fixed-effects meta-analysis approach.
3. We used standardized mean difference analysis to walk toward statistical homogeneity, because each study has a different number of observations within different periods.
4. As part of fixed-effect and standardized mean difference analysis, we applied the inverse variance statistical method by considering a 95% of confidence interval (CI).

According to Sect. 10.10.2 of Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2019), the simple usage of χ^2 to determine heterogeneity has low power in the typical situation of a meta-analysis when studies have a small sample size or are few. Instead, we adopt the I^2 approach to measure heterogeneity for better precision.

We prepared the synthesis results within a forest plot. The forest plot determines important synthesis parameters by including studies via p value and the I^2 heterogeneity measurement. We applied the Cochrane Review Manager 5.4.1 tool (RevMan) to generate funnel and forest plots. RevMan includes RoB 1.0 assessment. However, we replaced it with the updated RoB 2.0 for risk-of-bias analysis (Collaboration 2021).

We built a funnel plot to detect possible publication bias during the analysis of the results (Sects. 13.3.5.2 and 13.3.5.3 of Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2019)). We also linked the included studies to the research questions framework (both general and specific questions). For all studies that are missing outcomes, we proceeded with instructions from Sects. 10.12.2 and 10.12.3 of the Cochrane Handbook for

Table 3 All databases used for this systematic review

Database	Records retrieved
The ACM Guide to Computing Literature	638
Biomed Central Journals	61
BioRxiv + MedRxiv	93
BMJ Journals	25
Computers and Applied Sciences (EBSCOhost)	965
IEEE Xplore	16
JMIR Medical Informatics	168
Medline Daily Updates (Ovid)	165
Nature	114
Oxford Academic	119
PubMed PMC	1795
Sage Journals Online	370
ScienceDirect (Elsevier)	1289
Scopus (Elsevier)	21
SpringerLink	458
Web of Science	17
Wiley Online Library	717
Total	**7031**

Systematic Reviews of Interventions (Higgins et al. 2019) and did not exclude the studies. Instead, the forest and funnel plots discarded the studies with missing numerical outcomes.

Since the meta-analysis resulted in fixed effects and the heterogeneity coefficient is $I^2 = 0$. We agreed that we did not need to perform further meta-regression, sensitive, and other subgroup analyses, as they did not influence the final results of this systematic review.

3 Results

We now present the detailed results after applying the methods described in the previous sections. All results follow the PRISMA 2020 checklist (Higgins et al. 2019) and PRISMA protocol (Moher et al. 2015).

3.1 PRISMA workflow

The search strings initially returned 7031 records in database searching (see Table 3). We organized the results by following the PRISMA workflow according to Fig. 3.
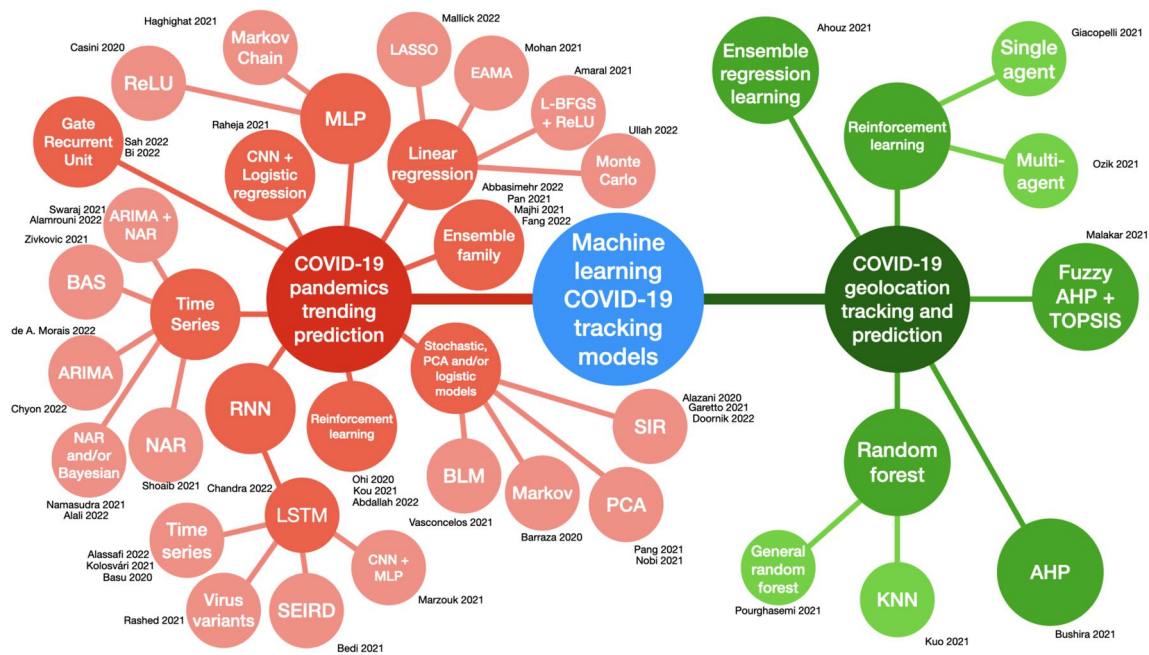
Fig. 3 PRISMA 2020 workflow
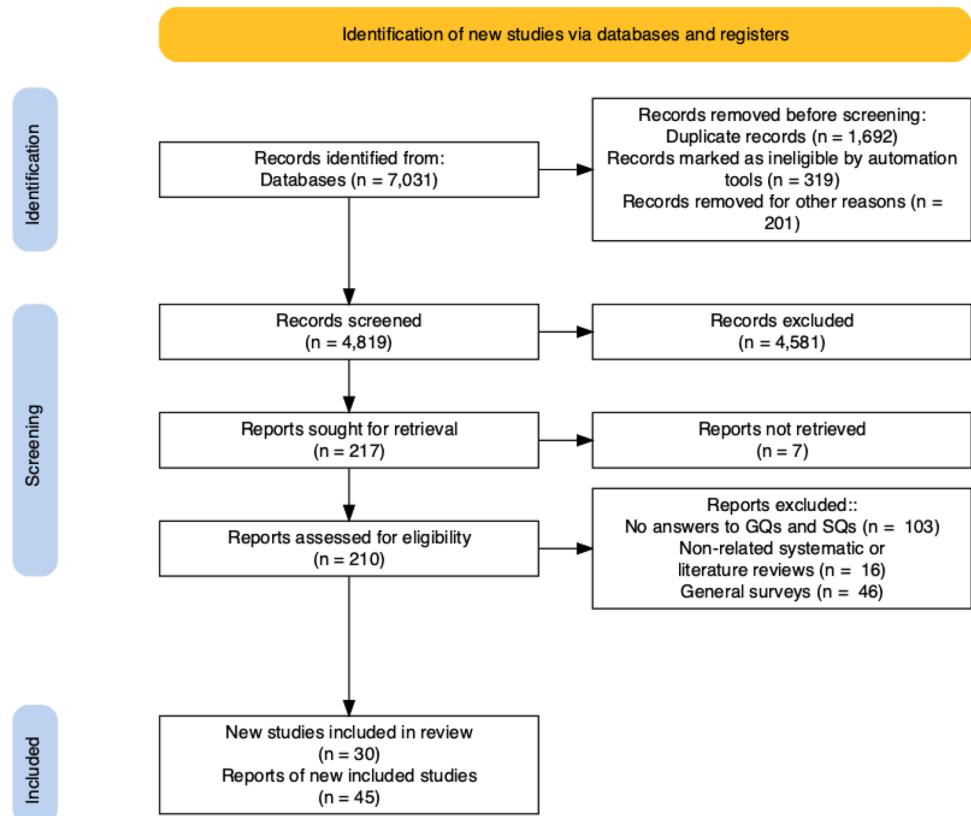
Fig. 4 Taxonomy of the studies included in this systematic review

After removing duplicated records, we marked records as ineligible by automation and excluded records for other reasons. We screened 4,819 records for seeking and based on the abstracts and keywords, and we excluded 4,581 records else. Then, from the resultant 217 records available for retrieval, we excluded seven records we could not retrieve.

We then reviewed 210 full-text documents and reached 45 included studies, grouped into 30 new unique studies. Finally, we performed meta-analyses of the results.

3.2 Taxonomy of the included studies

We obtained the 30 new studies from the 45 reports by performing a detailed taxonomy by dividing them into two main clusters, as detailed in Sect. 2.9.

Figure 4 describes the entire taxonomic organization of all studies in this systematic review.

1. **Domain 1: COVID-19 pandemics trending prediction: 38 studies**

   (a) Stochastic, principal component analysis (PCA), and logistic models: 7 studies
   (b) Reinforcement learning (RL): 3 studies
   (c) Recurrent neural networks (RNN) with long short-term memory (LSTM) architecture: 7 studies
   (d) Time-series: 8 studies
   (e) Gate recurrent unit (GRU): 2 studies
   (f) Convolutional neural network (CNN) with logistic regression: 1 study

   (g) Multilayer perceptron (MLP) neural network: 2 studies
   (h) Linear regression: 4 studies
   (i) Ensemble classifiers: 4 studies.

2. **Domain 2: COVID-19 geolocation tracking and prediction: 7 studies**

   (a) Ensemble regression learning: 1 study
   (b) RL: 2 studies
   (c) Fuzzy analytic hierarchy process (AHP) with Fuzzy technique for order of preference by similarity to ideal solution (TOPSIS): 1 study
   (d) AHP: 1 study
   (e) Random forest: 2 studies.

3.3 Answers to the research questions

The included studies addressed all research questions proposed by this paper, with different densities. We show the relationship between the included studies and the addressed questions in Table 4 and discuss them later in this section.

We have the following density of research questions answered:

- **GQ1**: 29 studies (96%)
- **GQ2**: 20 studies (69%)
- **SQ1**: 23 studies (84%)
- **SQ2**: 1 study (2%)
- **SQ3**: 9 studies (53%)

Table 4 Answers to research questions

Author, year	Country	Proposed methodology	Questions
Abbasimehr et al. (2022)	Iran, USA, Italy	Use time-series augmentation and Ensemble Bagging ML model to predict the incidence of COVID-19.	GQ1, GQ2, SQ1, SQ3
Abdallah et al. (2022)	Tunisia	Propose a new strategy based on reinforcement learning (RL) to study the dynamics of COVID-19 evolution	GQ1, GQ2, SQ1, SQ3
Ahouz and Golabpour (2021)	Iran	Predict the incidence of COVID-19 within two weeks to better manage the disease using the Least-Square Boosting Classification algorithm.	GQ1, SQ1, SQ4
Alali et al. (2022)	Saudi Arabia	Present an efficient machine learning method to forecast future trends of COVID-19 spread through Bayesian optimization	GQ1, GQ2, SQ1, SQ3
Alamrouni et al. (2022)	Saudi Arabia, Turkey, Nigeria, Cyprus	Predict the cumulative COVID-19 cases through ARIMA and ensemble ML models	GQ1, GQ2, SQ1, SQ3
Alassafi et al. (2022)	Saudi Arabia	Predict the cumulative COVID-19 positive cases by comparing RNN and LSTM neural networks' efficiency	GQ1, GQ2, SQ3
Alanazi et al. (2020)	Saudi Arabia, Canada	Model for predicting COVID-19 using the SIR and ML for innovative health care.	GQ1, SQ3, SQ4
Amaral et al. (2021)	Brazil	A data-driven SIR model whose parameters are fully calibrated by temporal functions learned from individual regressors and trained on different data sources.	GQ1, SQ1
Barraza et al. (2020)	Argentina	Model the epidemic spread as births in a population, where every birth corresponds to a new infection case. The paper introduces a new Markovian stochastic model that can be described as a non-homogeneous Pure Birth process.	GQ1, GQ2
Basu and Campbell (2020)	USA	A prediction long short-term memory (LSTM) based model that can be trained on more than four months of cumulative COVID-19 cases and deaths.	GQ1, GQ2, SQ1
Bedi et al. (2021)	India	A modified SEIRD model is based on the SEIRD model coupled with a deep learning-based long short-term memory (LSTM) model for COVID-19 case prediction trending.	GQ1, SQ1, SQ3, SQ4
Bi et al. (2022)	USA	Present a GRU-based hybrid model to predict the spread of COVID-19 across USA	GQ1, GQ2, SQ1, SQ3
Bushira and Ongala (2021)	Namibia	Using geospatial technologies, identify and map COVID-19 risk zones and model future COVID-19 responses in Namibia.	GQ1, GQ2, SQ1, SQ4
Casini and Roccetti (2020)	Italy	Present three computational models of increasing complexity (linear, negative binomial regression, and cognitive) to identify the one that better correlates the relationship between Italian tourist flows during the summer of 2020 and the resurgence of COVID-19 cases across the country.	GQ1, SQ1

Table 4 (continued)

Author, year	Country	Proposed methodology	Questions
Chandra et al. (2022)	Australia, India	Employ LSTM models to forecast the spread of COVID-19 infections among selected states in India.	GQ1, GQ2, SQ1, SQ3
Chyon et al. (2022)	Bangladesh	Predict the spread of COVID-19 infections by employing ARIMA and time-series machine learning.	GQ1, SQ1, SQ3
de Araújo Morais and da Silva Gomes (2022)	Brazil	Employ ARIMA with neural network models to forecast the spread of COVID-19 infections	GQ1, SQ1, SQ3
Doornik et al. (2022)	UK	Build a short-term forecasting ML model for the Coronavirus pandemic	GQ1, GQ2, SQ3
Fang et al. (2022)	China	Propose a data-driven XGBoost model for the prediction of COVID-19	GQ1, GQ2, SQ1, SQ3
Garetto et al. (2021)	Italy	Compared to traditional compartmental models, a time-modulated Hawkes process to model the spread of COVID-19.	GQ1, GQ2
Giacopelli (2021)	Italy	A full-scale individual-based model of the COVID-19 outbreak to test various scenarios about the pandemic and achieve novel performance metrics.	GQ1, SQ1, SQ4
Haghighat (2021)	Iran	A combined multilayer perceptron (MLP) neural network and Markov chain (MC) model to predict two indicators of the number of discharged and death cases according to their relationship with the number of hospitalized patients.	GQ1, GQ2, SQ1, SQ3
Kolozsvári et al. (2021)	Hungary	Use the official epidemiological data to forecast the epidemic curves (daily new cases) of COVID-19 using Artificial Intelligence (AI)-based Recurrent Neural Networks (RNNs), then compare and validate the predicted models with the observed data.	GQ1, GQ2, SQ1, SQ3
Kou et al. (2021)	China	A multi-scale agent-based model (MSABM) for disease spread between cities and in a town. The model investigates the infectious disease propagation between cities and within a city using the knowledge from person-to-person transmission.	GQ1, GQ2, SQ1
Kuo and Fu (2021)	USA	Develop a COVID-19 case predicting model based on county-level data with ML techniques, and the next-1-day (N1D), 4-day (N4D), and 7-day (N7D) averages of daily cases and cumulative cases would be used as a response respectively.	GQ1, GQ2, SQ1, SQ4
Majhi et al. (2021)	USA, India	Build predictive models that can predict the number of positive cases with higher accuracy. Regression-based, Decision tree-based, and Random forest-based models have been built on the data from China to validate India's sample.	GQ1, SQ3

Table 4 (continued)

Author, year	Country	Proposed methodology	Questions
Malakar (2021)	India	Model the COVID-19 vulnerability using an integrated fuzzy multi-criteria decision-making (MCDM) approach, namely fuzzy-analytical hierarchy process (AHP) and fuzzy-technique for order preference by similarity to ideal solution (TOPSIS) for West Bengal, India, through geographic information system (GIS).	GQ1, GQ2, SQ1, SQ4
Mallick et al. (2022)	India	Predict the COVID-19 trajectories of the infected population of Indian states using SEIR mathematical model and LASSO ML.	GQ1, SQ1, SQ3
Marzouk et al. (2021)	Egypt	Apply artificial intelligence-based models to predict the prevalence of the COVID-19 outbreak in Egypt. These models are long short-term memory networks (LSTM), convolutional neural networks, and multilayer perceptron neural networks.	GQ1, GQ2, SQ1, SQ3
Senthilkumar et al. (2021)	India, Australia, Pakistan, UK, USA, United Arab Emirates	A multimodel ML technique called ensemble learning, autoregressive, and moving regressive (EAMA). The multimodel forecasts COVID-19 parameters in the long term within India and globally.	GQ1, SQ1
Namasudra et al. (2021)	India	Present a novel Nonlinear Autoregressive (NAR) Neural Network Time-Series (NAR-NNTS) model to forecast COVID-19 cases. The authors train NAR-NNTS model with Scaled Conjugate Gradient (SCG), Levenberg–Marquardt (LM), and Bayesian Regularization (BR) training algorithms.	GQ1, GQ2, SQ1, SQ3
Nobi et al. (2021)	Bangladesh, South Korea	Apply principal component analysis (PCA) to the correlations for the changes of cumulative time-series of COVID-19 death and confirmed cases between different countries	GQ1, GQ2, SQ1, SQ3
Ohi et al. (2020)	Bangladesh, Saudi Arabia	Demonstrate what actions an agent (trained using reinforcement learning) may take in different possible pandemic scenarios depending on the spread of disease and economic factors.	GQ1, SQ1
Ozik et al. (2021)	USA, France	Present CityCOVID, which is a detailed agent-based model that enables the creation of efficient, urban-scale MPI-distributed agent-based models (ABMs).	GQ1, GQ2, SQ1, SQ3, SQ4
Pan et al. (2021)	Singapore, Taiwan, USA, Poland	A spatiotemporal analysis framework under the combination of an ensemble model (random forest regression) and a multi-objective optimization algorithm (NSGA-II).	GQ1, GQ2, SQ1
Pang et al. (2021)	China	Collect five indicators of the data of COVID-19 and its loss in Hubei Province. By applying PCA, the authors established the disaster loss index model of COVID-19 in Hubei Province.	GQ1, SQ1

Table 4 (continued)

Author, year	Country	Proposed methodology	Questions
Pourghasemi et al. (2020)	Iran, USA	Predicting mortality trends was built using regression modeling, spatial modeling, and risk mapping. The authors did the change detection using the random forest (RF) ML technique and the modeled risk map validation.	GQ1, GQ2, SQ1, SQ3, SQ4
Raheja et al. (2021)	India, UK	Present a novel approach (ML diffusion) prediction model for predicting the number of Coronavirus cases in four countries: India, France, China, and Nepal.	GQ1, GQ2, SQ1
Rashed and Hirata (2021)	Japan, Egypt	Analyze daily positive cases (DPC) data using a ML model to understand the effect of new viral variants on morbidity rates.	GQ2, SQ2
Sah et al. (2022)	Turkey, India, Saudi Arabia	Forecast COVID-19 Pandemic by combining ARIMA and LSTM-GRU Models in India	GQ1, SQ1, SQ3
Shoaib et al. (2021)	Taiwan, Pakistan, Thailand	Use the innovative design of the NAR-RBFs neural network paradigms designed to construct the SITR epidemic differential equation (DE) model to ascertain the different features of the spread of COVID-19.	GQ1, GQ2, SQ1
Swaraj et al. (2021)	India, Brazil	Present an ensemble model that integrates an autoregressive moving average model (ARIMA), and a nonlinear autoregressive neural network (NAR) to predict future cases of COVID-19.	GQ1, GQ2, SQ1
Ullah et al. (2022)	USA, Canada	Apply nonlinear modal regression and Monte Carlo algorithm to predict COVID-19 confirmed cases.	GQ2, SQ1, SQ3
Vasconcelos et al. (2021)	Brazil	A generalized logistic growth model with time-dependent parameters to describe the fatality curves of the COVID-19 disease. Such an approach applies to several countries that exhibit multiple waves of infections.	GQ1, GQ2, SQ1
Zivkovic et al. (2021)	Serbia, India, Vietnam, Turkey	Improve the current time-series prediction (forecasting) algorithms based on hybrid ML and nature-inspired algorithms to help predict future COVID-19 cases.	GQ1, GQ2, SQ1

- **SQ4**: 9 studies (20%).

Notice that SQ2 returned only one study (Rashed and Hirata 2021) from the search strategy executed between January 2020 and July 2022. One possible reason is that SARS-CoV-2 virus variants' impact acknowledgment began around the third quarter of 2020, and that would explain the lack of ML models focusing on virus variants. However, we need to consider the virus variants prediction and spreading during epidemiological studies as essential variables in building prediction models.

Some included studies have quick access to real-time COVID-19 confirmed cases data provided by their country governments or reliable institutions, facilitating predicting models more quickly. For instance, India's studies frequently cited the same data provided by government authorities. USA studies relied on Johns Hopkins COVID-19 Map[1]. Other studies frequently use official data from World Health Organization (WHO) databases, which may depend upon countries' slower case notifications. This scenario may stimulate some countries to step ahead in producing more studies about prediction models, while others do not have the same advantage.

---

[1] https://coronavirus.jhu.edu/map.html.

3.4 Risk-of-bias results

We assessed the risk of bias using RoB 2.0 tool (Higgins et al. 2019; Collaboration 2021) (see Sect. 2.7 for the domains assessed in this systematic review).

In terms of the overall risk of bias, the following studies presented high risk (5 of 45 included studies) (Kuo and Fu 2021; Pan et al. 2021; Pang et al. 2021; Shoaib et al. 2021; Zivkovic et al. 2021).

The following studies presented some level of concern (13 of 45 included studies) (Abdallah et al. 2022; Alamrouni et al. 2022; Bi et al. 2022; Chyon et al. 2022; Majhi et al. 2021; Senthilkumar et al. 2021; Nobi et al. 2021; Ohi et al. 2020; Ozik et al. 2021; Raheja et al. 2021; Sah et al. 2022; Ullah et al. 2022; Vasconcelos et al. 2021).

We present the risk of bias for each study in Fig. 5. We used the RobVis R tool to generate the risk of bias reports (McGuinness 2019).

We applied RoB 2.0 Tool over the two branches of our taxonomic analyses. That means we consider the risk of bias analysis over the context of each taxonomic context.

The studies from Kou et al. (2021) and Alanazi et al. (2020) show zero observations in the forest plot. However, they present outcome data according to their study contexts, which is why we rated their risk of bias low. The study from Haghighat 2021 (Haghighat 2021) shows zero observations in the forest plot, because it focuses on three categories only: hospitalized, deaths, and discharged COVID-19-positive patients. The total of cases is not the simple sum of such classes. However, the risk of bias is low, because the study shows the outcomes as expected (Fig. 6).

3.5 Overall effect size results of the individual studies

We present the forest plot in Fig. 7 by including all the meta-analysis results from each included study in this systematic review.

We need the forest plot analysis to determine the accuracy of the ML prediction models to represent actual cumulative COVID-19 cases. Therefore, the more accurate the models are, the less risky they represent for public health authorities' adoption.

The goal of a forest plot is to help physicians determine drug or treatment effectiveness within a population. Physicians are more interested in observing how many studies are far from the *line of no effect* and not passing through the *diamond* extension within the forest plot. Anything different from this approach means no statistical significance for physicians.

In our systematic review, we explore the exact opposite effect and interpretation. The studies closer to the *line of no effect* mean that they simulate real COVID-19 case scenarios

Fig. 5 Risk of bias from RoB 2.0 tool for each included study

more accurately. The same observation approach is valid for studies crossing the forest plot *diamond*. That is the statistical significance we aim for in this systematic review.

On the other hand, if actual case data reveal as too favored in the forest plot, it may indicate a risk of bias on missing or unclear outcomes presented by the study. The scenario is analyzed later in this review.

From Fig. 7, three case prediction studies are far from the *line of no effect*: Basu and Campbell (2020); Casini and Roccetti (2020) and (Kolozsvári et al. 2021). For the real case's side of the forest plot, only (Namasudra et al. 2021) is a little far from the *line of no effect*.

3.6 Results of syntheses

The forest plot presented heterogeneity $I^2 = 0$ for fixed and random effect studies. We chose the fixed-effects analysis for the reasons explained in Sect. 2.9.

Another critical aspect to observe in the forest plot is the summary of results: SMD=0.18, with 95% CI limits between [0.01, 0.35]. Since the interval does not cross the *line of no effect*, it means that the difference found between the two groups was statistically significant. As we used SMD calculation, the overall impact returns Z statistic and the corresponding p value: $Z = 2.02$ and p value=0.04. Furthermore, since p value≤ 0.05, the summary result is inside the statistical significance threshold established by medical and psychological research (Sect. 15.3.2 of Cochrane handbook (Higgins et al. 2019)). In this specific case, the overall effect means that the included ML prediction studies are significantly valid in simulating the trend of the real cumulative COVID-19 infection cases.

The forest plot (Fig. 7) presents zero observations for most geolocation studies due to focusing on the pandemic prediction simulation, not geolocation. However, the zero observations are discarded by RevMan 5.4.1 and do not influence the results of forest plot statistical analyses. The zero observations reflect the risk of bias within domain D3

(bias due to missing outcome data). We indicate it as *No information* statement for missing outcomes.

3.7 Risk of bias due to missing results

Based on the forest plot from Fig. 7, we flagged the following studies with *No information* in domain D3 within the risk-of-bias analysis. The referred studies do not match the forest plot proposed synthesis (predicted vs. the actual number of cumulative COVID-19-positive cases) (Abdallah et al. 2022; Alali et al. 2022; Alamrouni et al. 2022; Alanazi et al. 2020; Bushira and Ongala 2021; Bi et al. 2022; Chandra et al. 2022; Chyon et al. 2022; de Araújo Morais and da Silva Gomes 2022; Haghighat 2021; Kou et al. 2021; Kuo and Fu 2021; Malakar 2021; Mallick et al. 2022; Nobi et al. 2021; Ohi et al. 2020; Ozik et al. 2021; Pan et al. 2021; Pang et al. 2021; Rashed and Hirata 2021; Sah et al. 2022; Shoaib et al. 2021; Ullah et al. 2022).

All the above studies present zero number of patients and the impossibility of SMD calculations. However, they do not alter the overall pooled effects calculated within the forest plot. Because of the reasons already detailed in Sect. 2.7, we did not exclude these studies, as they provide essential methodologies for this systematic review.

3.8 Risk of publication bias

We then grouped all studies that produced results to generate the funnel plot (Fig. 8) and analyzed aspects related to publication bias.

The funnel plot places most of the studies closer to the funnel's center, with only one study (Basu and Campbell 2020) as an outlier within 95% CI. This aspect of the funnel plot indicates a low probability of publication bias. In other words, the interpretation is that the ML prediction studies can simulate real-case scenarios with a high chance of certainty, reinforcing the p value=0.04 calculated by the forest plot.
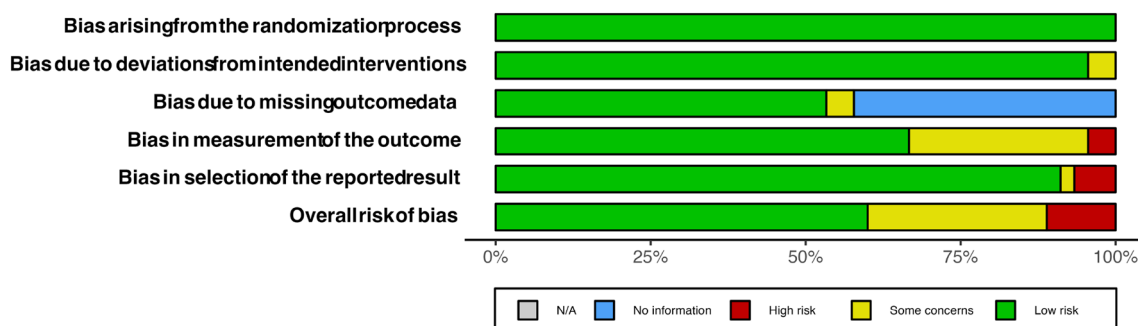

Fig. 6 Risk-of-bias summary from RoB 2.0 tool

Study or Subgroup	Actual Cases			Predicted Cases			Weight	Std. Mean Difference IV, Fixed, 95% CI
	Mean	SD	Total	Mean	SD	Total		
Abbasimehr 2022	2,949	60.210741	7	2,943.857143	57.0889031	7	2.7%	0.08 [-0.97, 1.13]
Abdallah 2022	0	0	0	0	0	0		Not estimable
Ahouz 2021	81,001.28571	8,824.995	14	79,761.3571	7,491.00007	14	5.4%	0.15 [-0.59, 0.89]
Alali 2022	0	0	0	0	0	0		Not estimable
Alamrouni 2022	0	0	0	0	0	0		Not estimable
Alanazi 2020	0	0	0	0	0	0		Not estimable
Alassafi 2022	1,169.28571	219.704889	7	1,171.42857	200.035711	7	2.7%	-0.01 [-1.06, 1.04]
Amaral 2021	563,363.636	48,378.2445	11	548,818.1818	50,459.52	11	4.2%	0.28 [-0.56, 1.12]
Barraza 2020	504,459.875	662,060.491	8	515,825	681,694.6	8	3.1%	-0.02 [-1.00, 0.96]
Basu 2020	2,256,462.96	228,550.025	27	2,080,693	126,149.5	27	9.3%	0.94 [0.37, 1.50]
Bedi 2021	12,748.3333	22,522.1594	15	11,909.33333	21,482.35	15	5.8%	0.04 [-0.68, 0.75]
Bi 2022	0	0	0	0	0	0		Not estimable
Bushira 2021	0	0	0	0	0	0		Not estimable
Casini 2020	1,796.95	1,710.96016	20	1,105.9	1,106.138	20	7.5%	0.47 [-0.16, 1.10]
Chandra 2022	0	0	0	0	0	0		Not estimable
Chyon 2022	0	0	0	0	0	0		Not estimable
de Araújo Morais 2022	0	0	0	0	0	0		Not estimable
Doornik 2022	33,000	12,328.828	8	33,000	12,328.828	8	3.1%	0.00 [-0.98, 0.98]
Fang 2022	96,300	50,214.0419	5	96,400	50,346.7973	5	1.9%	-0.00 [-1.24, 1.24]
Garetto 2021	131,333.625	102,973.384	16	130,500	101,407.4	16	6.1%	0.01 [-0.69, 0.70]
Giacopelli 2021	1,770	1,365.46695	5	1,550	1,029.563	5	1.9%	0.16 [-1.08, 1.41]
Haghighat 2021	0	0	0	0	0	0		Not estimable
Kolozsvári 2021	3,433.33333	479.267117	12	3,070.83333	460.9567	12	4.3%	0.74 [-0.09, 1.58]
Kou 2021	0	0	0	0	0	0		Not estimable
Kuo 2021	0	0	0	0	0	0		Not estimable
Majhi 2020	11,366	11,208.0795	8	12,650	14,533.41	8	3.1%	-0.09 [-1.07, 0.89]
Malakar 2021	0	0	0	0	0	0		Not estimable
Mallick 2022	0	0	0	0	0	0		Not estimable
Marzouk 2021	41,795	44,222.2551	7	41,685.71429	44,679.88	7	2.7%	0.00 [-1.05, 1.05]
Mohan 2021	3,233,344.2	2,060,955.89	10	3,140,000	1,349,238.47	10	3.8%	0.05 [-0.83, 0.93]
Namasudra 2021	7,730,764.89	2,135,085.37	19	8,230,341.105	2,867,036.6	19	7.3%	-0.19 [-0.83, 0.44]
Nobi 2021	0	0	0	0	0	0		Not estimable
Ohi 2020	0	0	0	0	0	0		Not estimable
Ozik 2021	0	0	0	0	0	0		Not estimable
Pan 2021	0	0	0	0	0	0		Not estimable
Pang 2021	0	0	0	0	0	0		Not estimable
Pourghasemi 2020	4,723.33333	2,792.75846	14	4,773.33333	2,769.66132	14	5.4%	-0.02 [-0.76, 0.72]
Raheja 2021	3,834,972	516,255,958	4	3,834,972	516,255,958	4	1.5%	0.00 [-1.39, 1.39]
Rashed 2021	0	0	0	0	0	0		Not estimable
Sah 2022	0	0	0	0	0	0		Not estimable
Shoaib 2021	0	0	0	0	0	0		Not estimable
Swaraj 2021	3,622.27273	322.423352	11	3,558.636364	283.302761	11	4.2%	0.20 [-0.64, 1.04]
Ullah 2022	0	0	0	0	0	0		Not estimable
Vasconcelos 2021	130,875	111,517.374	8	130,875	111,517.374	8	3.1%	0.00 [-0.98, 0.98]
Zivkovic 2021	25,337.2414	20,662.5405	29	24,597.27586	22,009.7687	29	11.1%	0.03 [-0.48, 0.55]
Total (95% CI)			265			265	100.0%	0.18 [0.01, 0.35]

Heterogeneity: Chi² = 13.09, df = 21 (P = 0.91); I² = 0%
Test for overall effect: Z = 2.02 (P = 0.04)

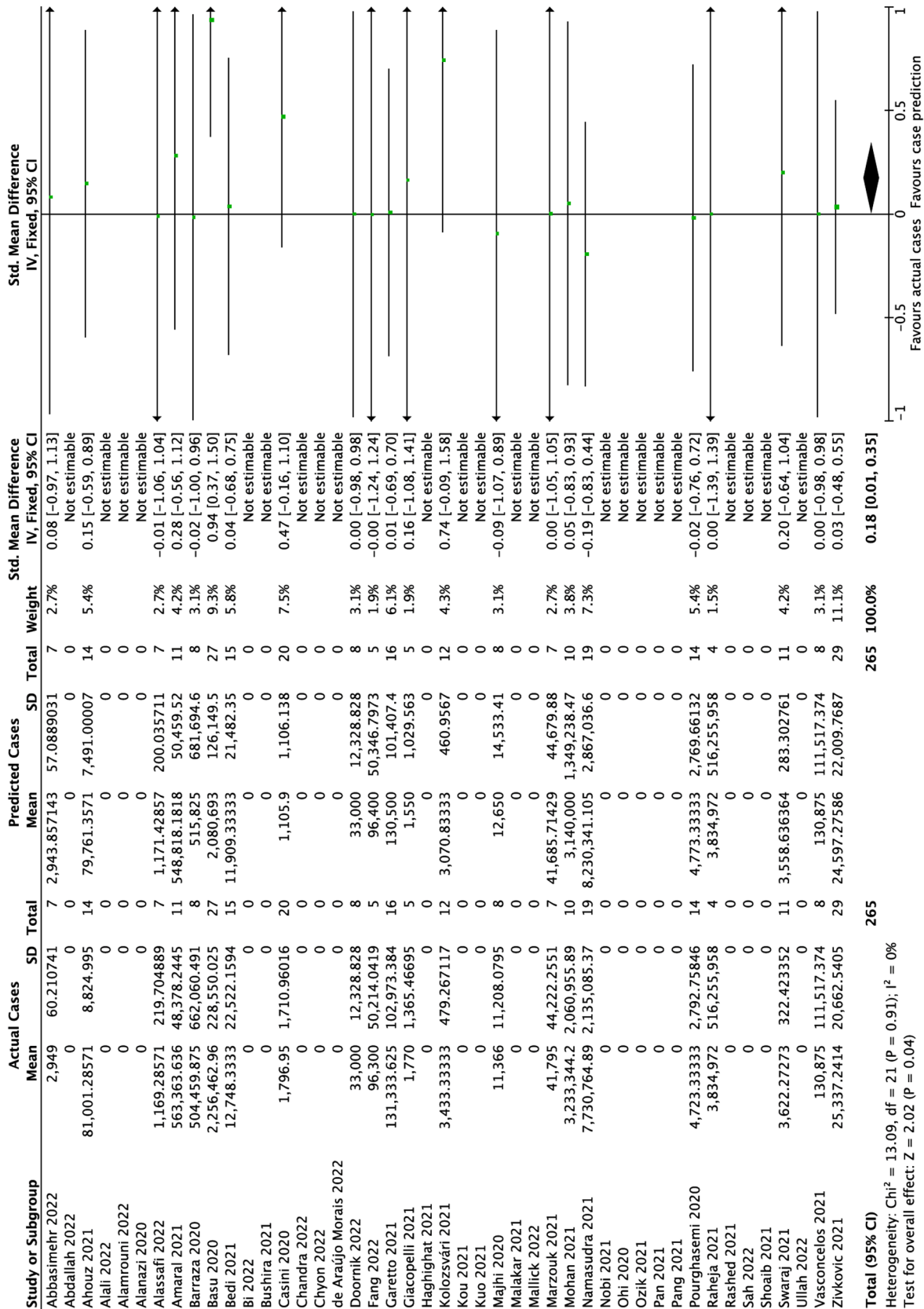Favours actual cases   Favours case prediction

Fig. 7 Forest plot for effect sizes. The figure displays the meta-analysis consisting of the summary statistics [standardized mean (SMD), standard deviation (SD), sample size, and weight proportion] for comparison between predicted cases by ML model vs. actual COVID-19 cases. It also displays the mean difference and its 95% confidence interval (CI) for the continuous outcome

Fig. 8 Funnel plot based on non-zero studies from forest plot (Fig. 7). The *x*-axis shows the SMD, while the *y*-axis shows the standard error (SE)
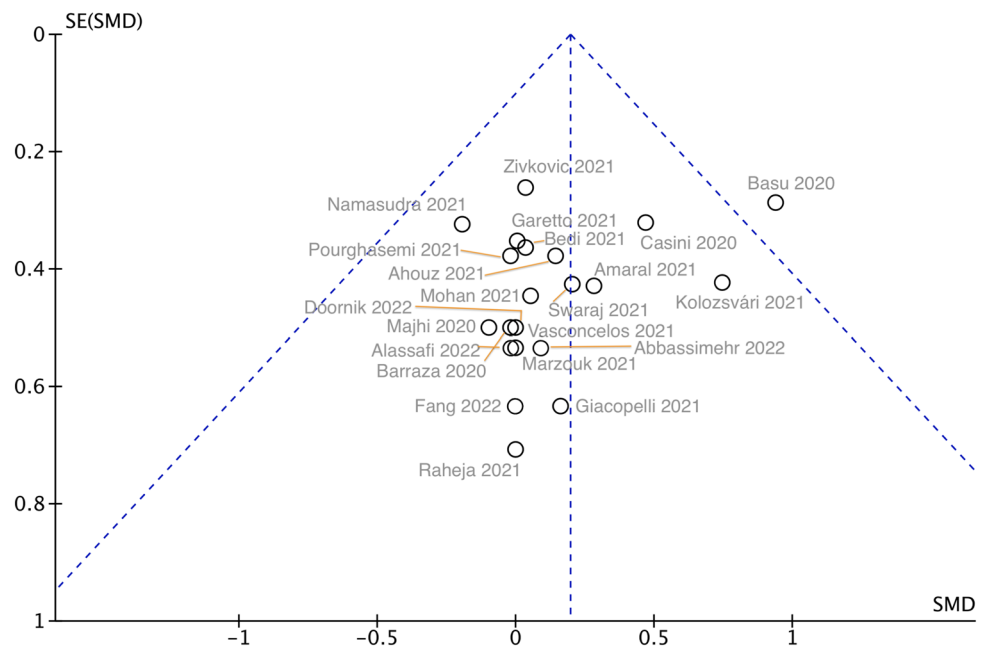
The GRADE summary of findings also includes the publication bias when presenting the pooled study results for the certainty of evidence.

3.9 Certainty of evidence

We presented the certainty of evidence assessment through the GRADE summary of finding tables (Schunemann 2013) divided into the two domains defined in Sect. 2.6. The GRADE summary of findings tables was created and assessed based on the taxonomy defined in Sect. 3.2.

A summary of findings for domains 1 and 2 revealed the following pooled studies and corresponding levels of certainty. Further details are available in GRADE summary of findings tables in Figs. 9 and 10.

- **Domain 1: COVID-19 pandemics' trending prediction**

  - **Moderate level of certainty**: Linear regression, LSTM RNN, stochastic and PCA logistic models, and ensemble classifiers' family
  - **Low level of certainty**: CNN with logistic regression, time-series neural network, MLP, and GRU
  - **Very low level of certainty**: Reinforcement learning.

- **Domain 2: COVID-19 geolocation tracking and prediction**

  - **High level of certainty**: Ensemble regression learning
  - **Low level of certainty**: Fuzzy AHP and TOPSIS, AHP, and random forest

- **Very low level of certainty**: Reinforcement learning.

3.10 Limitations

The first limitation of this systematic review is that each study has access to heterogeneous databases, and there are no common databases where we could promote studies crossovers. The forest plot (Fig. 7) statistically normalizes the data to calculate a common SMD. However, we consider that studies crossover to common databases would bring a plus significance to this systematic review.

Some studies provided outcomes in the format of low-resolution graphics, which turned the task of numerical reading more difficult. In these cases, authors should provide the supplementary material with the numerical data from where they built the outcome graphics or a high-resolution graphics version. We needed to read the data more closely to avoid bias and take the best approximated numerical values.

As we mentioned in the risk-of-bias Sect. 3.7, some studies brought innovative ML methods to address COVID-19 prediction, but presented no results or numerical and graphical outcomes of the prediction results. Although Sect. 10.12.2 of Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2019) allows replacement data fulfillment for cases of lack of numerical outcomes, we did not adopt this approach to avoid any biased introduction to this systematic review. Instead, we decided to rely on the graphics data regardless they were not accurate. To preserve the original level of

№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	machine learning prediction	real cumulative cases	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
Linear regression												
4	randomised trials	serious[a]	not serious[a]	not serious	not serious[a]	none	21	21	-	SMD 0.17 SD higher (0.43 lower to 0.78 higher)	⊕⊕⊕◯ Moderate	CRITICAL
Convolutional neural network with Logistic regression												
1	randomised trials	not serious	not serious	not serious	serious[b]	publication bias strongly suspected[b]	4	4	-	SMD 0 SD (1.39 lower to 1.39 higher)	⊕⊕◯◯ Low	CRITICAL
Long Short-Term Memory (LSTM) recurrent neural networks (RNN)												
7	randomised trials	serious[c]	not serious	not serious	not serious[c,d]	none	68	68	-	SMD 0.49 SD higher (0.14 higher to 0.83 higher)	⊕⊕⊕◯ Moderate	CRITICAL
Stochastic, PCA and logistic models												
7	randomised trials	not serious	not serious	not serious	serious[e,f]	none	40	40	-	SMD 0 SD (0.44 lower to 0.44 higher)	⊕⊕⊕◯ Moderate	IMPORTANT
Time series models												
8	randomised trials	not serious	serious[g]	not serious	serious[g]	none	59	59	-	SMD 0.01 SD lower (0.37 lower to 0.35 higher)	⊕⊕◯◯ Low	CRITICAL
Multilayer perceptron												
2	randomised trials	serious[h]	not serious	not serious	serious[h]	none	20	20	-	SMD 0.47 SD higher (0.16 lower to 1.1 higher)	⊕⊕◯◯ Low	CRITICAL
Ensemble family												
4	randomised trials	not serious	not serious	not serious	serious[i]	none	20	20	-	SMD 0.01 SD lower (0.63 lower to 0.61 higher)	⊕⊕⊕◯ Moderate	IMPORTANT
Reinforcement learning												
3	randomised trials	serious[j]	not serious	not serious	serious[j]	publication bias strongly suspected[j]	0	0	-	SMD 0 SD (0 to 0)	⊕◯◯◯ Very low	NOT IMPORTANT
Gate recurrent unit												
2	randomised trials	not serious	not serious	not serious	serious[k]	publication bias strongly suspected[l]	0	0	-	0 (0 to 0)	⊕⊕◯◯ Low	NOT IMPORTANT

Fig. 9 Question for domain 1 GRADE assessment: ML prediction compared to real cumulative cases for tracking COVID-19 pandemics' future trending. Explanations: (**a**) Mohan et al. 2021 study should compare real and predicted cases; (**b**) Raheja et al. 2021 study presents an abnormal standard error and increases the risk of publication bias; (**c**) Marzouk et al. 2021 study should compare actual and predicted cases; (**d**) Basu et al. 2020 appear as an outlier in the funnel plot, indicating possible imprecision; (**e**) Vasconcelos et al. 2021 bring results, but presenting a low-graph resolution, which may cause imprecisions; (**f**) Pang et al. 2021 study should bring the predicted vs. real cases, as this is the goal of this study; (**g**) Shoaib et al. 2021 bring no outcomes; (**h**) Haghighat 2021 brings no outcomes; (**i**) Pan et al. 2020 graphs with predicted vs. real cases have low resolution, affecting exact values from the study; (**j**) Ohi et al. 2020 and Kou et al. 2021 do not bring results about predicted vs. real cumulative cases; (**k, l**) Sah 2022 and Bi 2022 should bring the predicted vs. real cases, as this is the goal of this study. There is only continuity from actual to predicted cases (Senthilkumar et al. 2021; Raheja et al. 2021; Marzouk et al. 2021; Basu and Campbell 2020; Vasconcelos et al. 2021; Pang et al. 2021; Shoaib et al. 2021; Haghighat 2021; Pan et al. 2021; Ohi et al. 2020; Kou et al. 2021; Sah et al. 2022; Bi et al. 2022)

Certainty assessment							Number of patients		Effect		Certainty	Importance
Number of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	machine learning prediction	real geographic location	Relative (95% CI)	Absolute (95% CI)		
Ensemble regression learning												
1	randomised trials	not serious	not serious	not serious	not serious	none	14	14	-	SMD **0.15 SD higher** (0.59 lower to 0.89 higher)	⊕⊕⊕⊕ High	CRITICAL
Reinforcement learning												
2	randomised trials	serious[a]	not serious	not serious	very serious[a]	none	5	5	-	SMD **0.16 SD higher** (1.08 lower to 1.41 higher)	⊕◯◯◯ Very low	CRITICAL
Fuzzy analytic hierarchical process (AHP) with Fuzzy technique for order of preference by similarity to ideal solution (TOPSIS)												
1	randomised trials	serious[b]	not serious	not serious	serious[b]	none			-	**0** (0 to 0)	⊕⊕◯◯ Low	CRITICAL
Analytic hierarchical process (AHP)												
1	randomised trials	serious[c]	not serious	not serious	serious[c]	none			-	**0** (0 to 0)	⊕⊕◯◯ Low	IMPORTANT
Random forest												
2	randomised trials	serious[d]	not serious	not serious	serious[d]	none	14	14	-	SMD **0.02 SD lower** (0.76 lower to 0.72 higher)	⊕⊕◯◯ Low	CRITICAL

Fig. 10 Question for domain 2 GRADE assessment: ML prediction compared to real geographic location for tracking COVID-19 pandemics critical geolocations: (**a**) Ozik et al. 2021 have no outcomes for cumulative COVID-19 cases' prediction. Giacopelli 2021 brings very few observations; (**b**) Malakar et al. 2021 bring imprecise outcomes for cumulated COVID-19 cases during modeling; (**c**) Bushira et al. 2021 have no outcomes for cumulative COVID-19 cases prediction; (**d**) Kuo et al. 2021 have no outcomes for cumulative COVID-19 cases' prediction (Ozik et al. 2021; Giacopelli 2021; Malakar 2021; Bushira and Ongala 2021; Kuo and Fu 2021)

certainty, we considered either no data or approximated data (due to low-resolution graphics in some studies).

4 Conclusion

This article presented a systematic review highlighting the contribution of applied computing to the health area. There is a significant effort to model, predict, and mitigate the COVID-19 pandemic across the population. We conducted the systematic review based on the PRISMA protocol, where we could conduct the full review of 45 studies out of initially 7031 screened studies. We could then answer the general and specific questions proposed in this systematic review based on the PICO approach.

Compared to other systematic reviews cited in this study ((Ghaderzadeh and Asadi 2021; Alyasseri et al. 2022; Syeda et al. 2021), our systematic review has the advantage of providing a straightforward scientific contribution to joining all ML pandemic evolution models into a single study and comparing which model can be the best fit depending upon the affected region criticality.

Using Cochrane methodology and GRADE assessment for a summary of findings table allowed us to dive deep into the studies and determine their significance to the scientific contribution that can act as a speedy response to minimize the health disasters caused by the COVID-19 pandemics.

We observed essential ML approaches to predict the next steps of the pandemics regarding COVID-19-positive trends and geolocation. By conducting risk of bias, levels of certainty, a summary of findings, and statistical analysis via the forest and funnel plot assessments for the 45 studies, we could determine the statistical significance of these studies to simulate the progress of the pandemics. Despite some study limitations found during this systematic review, our final results corroborate the possibility of using ML prediction to serve the healthcare authorities for decision-making and preventive actions toward saving lives. ML and healthcare can offer valuable options to respond to and combat current and future pandemics.

The healthcare authorities can take immediate advantage of using the included studies from this systematic review and implement any of the corresponding ML models in the format of ML pipeline graphs or Kubernetes[2] container resources already present in the cloud hyperscalers (like Google Cloud Platform[3], SAP Data Intelligence Cloud[4], Microsoft Azure[5] or Amazon Web Services[6]). The containers will provide endpoints where users can enter big data for training and testing.

Then, the trained ML container seamlessly becomes a production microservice available for final users.

5 Registration and protocol

This systematic review has not been registered in the PROSPERO database yet.

The citation of this systematic review and meta-analysis protocol is presented in Moher et al. (2015). We explain the details about differences in resultant statistical significance through this systematic review, but they do not cause any differences in the original PRISMA-P protocol.

Appendix A: Detailed search strings for each queried database

Per PRISMA 2020 checklist item 7 (Page et al. 2021), we present the individual search strings we used for each database below. Note that not all queries allow explicit periods. Due to this reason, publication periods will not explicitly appear in the search string; instead, we filter them after results retrieval. However, as described in Table 1, we always define the referred period to ensure the correct retrieved results. For search strings where January 2020-July 20222 is not explicit, we filtered into the corresponding period later, after the journal returns the result set.

1. **The ACM Guide to Computing Literature** *[[Abstract: covid-19] OR [Abstract: sars-cov-2]] AND [Abstract: pandemic scenario] AND [[Abstract: epidemiological model] OR [Abstract: predictive model]] AND [Abstract: machine learning] AND [Publication Date: (01/01/2020 TO 07/31/2022)]*
2. **Biomed Central Journals** *(COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning)*
3. **BioRxiv + MedRxiv** *(COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning)*

---

2 Kubernetes: https://kubernetes.io.

3 Google Cloud Platform: https://cloud.google.com.

4 SAP Data Intelligence Cloud: https://www.sap.com/dataintelligence.

5 Microsoft Azure: https://azure.microsoft.com.

6 Amazon Web Services: https://aws.amazon.com.

4. **EBSCOhost Computers and Applied Sciences** *AB pandemic scenario AND AB epidemiological model OR AB predictive model AND AB (machine learning or artificial intelligence) AND AB COVID-19 OR AB sars-cov-2*

5. **IEEE Xplore** *(COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning)*

6. **ElSevier ScienceDirect** *((COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning)*

7. **ElSevier Scopus** *TITLE-ABS-KEY((covid-19 OR sars-cov-2) AND (pandemic AND scenario) AND ((epidemiological AND model) OR (predictive AND model)) AND (machine AND learning)) AND PUBYEAR > 2020 AND PUBYEAR < 2022*

8. **JMIR Medica Informatics** *(COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning)*

9. **Ovid MEDLINE(R) Daily Update** *1. (COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning) {No Related Terms} 2. limit 1 to yr="2020 - 2022"*

10. **Oxford Academic (all journals)** *(COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning)*

11. **NCBI PubMed PMC** *Search (("COVID-19"[All Fields] OR "COVID-19"[MeSH Terms] OR "COVID-19 Vaccines"[All Fields] OR "COVID-19 Vaccines"[MeSH Terms] OR "COVID-19 serotherapy"[All Fields] OR "COVID-19 Nucleic Acid Testing"[All Fields] OR "covid-19 nucleic acid testing"[MeSH Terms] OR "COVID-19 Serological Testing"[All Fields] OR "covid-19 serological testing"[MeSH Terms] OR "COVID-19 Testing"[All Fields] OR "covid-19 testing"[MeSH Terms] OR "SARS-CoV-2"[All Fields] OR "sars-cov-2"[MeSH Terms] OR "Severe Acute Respiratory Syndrome Coronavirus 2"[All Fields] OR "NCOV"[All Fields] OR "2019 NCOV"[All Fields] OR (("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields] OR "COV"[All Fields]) AND 2020/01/01[PubDate] : 2022/07/31[PubDate])) OR ("sars-cov-2"[MeSH Terms] OR "sars-cov-2"[All Fields] OR "sars cov 2"[All Fields])) AND (("pandemics"[MeSH Terms] OR "pandemics"[All Fields] OR "pandemic"[All Fields]) AND scenario[All Fields]) AND ((("epidemiological"[All Fields] AND "models"[All Fields]) OR "epidemiological models"[All Fields] OR ("epidemiological"[All Fields] AND "model"[All Fields]) OR "epidemiological model"[All Fields]) OR (predictive[All Fields] AND model[All Fields])) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields]) AND ("2020/01/01"[PubDate] : "2022/07/31"[PubDate])*

12. **Sage Journals Online** *for [[All covid-19] OR [All sars-cov-2]] AND [All pandemic scenario] AND [[All epidemiological model] OR [All predictive model]] AND [All machine learning] within SAGE Open 2020-2022*

13. **Springer Nature Biomed Central Journals** *COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning)*

14. **Springer Nature Nature.com** *(COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning)*

15. **Springer Nature SpringerLink** *(COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning)*

16. **Web of Science** *ALL=((COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning))*

17. **Wiley Online Library** *(COVID-19 OR SARS-Cov-2) AND (pandemic scenario) AND ((epidemiological model) OR (predictive model)) AND (machine learning)*

Appendix B. List of acronyms used in this systematic review

This is the glossary of acronyms used in this systematic review:

- AGM - Agent-based model
- AHP - Analytic hierarchy process
- AI - Artificial intelligence
- ARIMA - Autoregressive integrated moving average
- BR - Bayesian regularization
- CNN - Convolutional neural network
- DP - Deep learning
- DPC - Daily positive cases
- DE - Differential equations
- EAMA - Ensemble learning, autoregressive, and moving regressive

- GIS - Geographical information system
- GQ - General question
- GRADE - Grading of recommendations assessment, development, and evaluation
- GRU - Gate recurrent unit
- LM - Levenberg–Marquardt
- LSTM - Long short-term memory
- MC - Markov chain
- MCDM - Multi-criteria decision-making
- MDT - Mendeley desktop tool
- ML - Machine learning
- MLP - Multilayer perceptron
- MM - Mathematical models
- MPI - Message passing interface
- MRM- Mendeley reference manager
- MSABM - Multi-scale agent-based model
- MWB - Mendeley web importer
- N1D - Next 1-day
- N4D - Next 4-day
- N7D - Next 7-day
- NAR - Nonlinear autoregressive
- NNTS - Neural network time-series
- NSGA-II - Non-dominated sorting genetic algorithm II
- PCA - Principal component analysis
- PICO - Participants, intervention, comparison, and outcomes
- PRISMA - Preferred reporting items for systematic reviews and meta-analyses
- PRISMA-P - Preferred reporting items for systematic reviews and meta-analyses' protocol
- RBF - Radial base function
- RF - Random forest
- RL - Reinforcement learning
- RNN - Recurrent neural network
- RoB - Risk of bias
- SCG - Scaled conjugate gradient
- SEIR - Susceptible, exposed, infectious, or recovered
- SEIRD - Susceptible, exposed, infectious, recovered, or deceased
- SIR - Susceptible, infectious, or recovered
- SITR - Susceptible, infected, treatment, recovered
- SMD - Standardized mean differences
- SQ - Specific question
- TOPSIS - Technique for order of preference by similarity to ideal solution
- WHO - World Health Organization.

Compliance with ethical standards

Competing interests 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.

References

Abbasimehr H, Paki R, Bahrini A (2022) A novel approach based on combining deep learning models with statistical methods for COVID-19 time series forecasting. Neural Comput Appl 34(4):3135–3149

Abdallah W, Kanzari D, Sallami D, Madani K, Ghedira K (2022) A deep reinforcement learning based decision-making approach for avoiding crowd situation within the case of COVID-19 pandemic. Comput Intell 38(2):416–437

Ahouz F, Golabpour A (2021) Predicting the incidence of COVID-19 using data mining. BMC Public Health 21(1):1–12

Alali Y, Harrou F, Sun Y (2022) A proficient approach to forecast COVID-19 spread via optimized dynamic machine learning models. Sci Rep 12(1):1–20

Alamrouni A, Aslanova F, Mati S, Maccido HS, Jibril AA, Usman A, Abba S (2022) Multi-regional modeling of cumulative COVID-19 cases integrated with environmental forest knowledge estimation: A deep learning ensemble approach. Int J Environ Res Public Health 19(2):738

Alanazi SA, Kamruzzaman M, Alruwaili M, Alshammari N, Alqahtani SA, Karime A (2020) Measuring and preventing COVID-19 using the SIR model and machine learning in smart health care. J Healthc Eng 2020:25

Alassafi MO, Jarrah M, Alotaibi R (2022) Time series predicting of COVID-19 based on deep learning. Neurocomputing 468:335–344

Alyasseri ZAA, Al-Betar MA, Doush IA, Awadallah MA, Abasi AK, Makhadmeh SN, Alomari OA, Abdulkareem KH, Adam A, Damasevicius R et al (2022) Review on COVID-19 diagnosis models based on machine learning and deep learning approaches. Expert Syst 39(3):12759

Amaral F, Casaca W, Oishi CM, Cuminato JA (2021) Towards providing effective data-driven responses to predict the COVID-19 in São Paulo and Brazil. Sensors 21(2):540

Barraza NR, Pena G, Moreno V (2020) A non-homogeneous Markov early epidemic growth dynamics model. Application to the SARS-CoV-2 pandemic. Chaos Solitons Fract 139:110297

Basu S, Campbell RH (2020) Going by the numbers: learning and modeling COVID-19 disease dynamics. Chaos Solitons Fract 138:110140

Bedi P, Dhiman S, Gole P, Gupta N, Jindal V (2021) Prediction of COVID-19 trend in India and its four worst-affected states using modified SEIRD and LSTM models. SN Comput Sci 2(3):1–24

Bi L, Fili M, Hu G (2022) COVID-19 forecasting and intervention planning using gated recurrent unit and evolutionary algorithm. Neural Comput Appl 2:1–19

Booth A, Noyes J, Flemming K, Moore G, Tunçalp Ö, Shakibazadeh E (2019) Formulating questions to explore complex interventions within qualitative evidence synthesis. BMJ Glob Health 4(Suppl 1):001107

Buhat CAH, Torres MC, Olave YH, Gavina MKA, Felix EFO, Gamilla GB, Verano KVB, Babierra AL, Rabajante JF (2021)

A mathematical model of COVID-19 transmission between frontliners and the general public. Netw Model Anal Health Inf Bioinf 10(1):1–12

Bushira KM, Ongala JO (2021) Modeling transmission dynamics and risk assessment for COVID-19 in Namibia using geospatial technologies. Trans Indian Natl Acad Eng 6(2):377–394

Casini L, Roccetti M (2020) A cross-regional analysis of the COVID-19 spread during the 2020 Italian vacation period: results from three computational models are compared. Sensors 20(24):7319

Chandra R, Jain A, Singh Chauhan D (2022) Deep learning via LSTM models for COVID-19 infection forecasting in India. PLoS ONE 17(1):0262708

Chyon FA, Suman MNH, Fahim MRI, Ahmmed MS (2022) Time series analysis and predicting COVID-19 affected patients by ARIMA model using machine learning. J Virol Methods 301:114433

Collaboration C, et al (2020) Review manager (RevMan)[computer program] version 5.4. Copenhagen: The Nordic Cochrane Centre

Collaboration C. et al (2021) Risk of bias 2 Cochrane review group starter pack. Cochrane Methods

de Araújo Morais LR, da Silva Gomes GS (2022) Forecasting daily COVID-19 cases in the world with a hybrid ARIMA and neural network model. Appl Soft Comput 126:109315

Doornik JA, Castle JL, Hendry DF (2022) Short-term forecasting of the Coronavirus pandemic. Int J Forecast 38(2):453–466. https://doi.org/10.1016/j.ijforecast.2020.09.003

Fang Z-G, Yang S-Q, Lv C-X, An S-Y, Wu W (2022) Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study. BMJ Open 12(7):056685

Fox M (2020) Rich nations have grabbed more than half the Coronavirus vaccine supply already, report finds. CNN . Accessed 2021-06-11

Gallagher J (2021) COVID: Is there a limit to how much worse variants can get? BBC UK . Accessed 2021-09-09

Garetto M, Leonardi E, Torrisi GL (2021) A time-modulated Hawkes process to model the spread of COVID-19 and the impact of countermeasures. Ann Rev Control 2:2

Gartlehner G, Affengruber L, Titscher V, Noel-Storr A, Dooley G, Ballarini N, König F (2020) Single-reviewer abstract screening missed 13 percent of relevant studies: a crowd-based, randomized controlled trial. J Clin Epidemiol 121:20–28

Ghaderzadeh M, Asadi F (2021) Deep learning in the detection and diagnosis of COVID-19 using radiology modalities: a systematic review. J Healthc Eng 2021:2

Giacopelli G et al (2021) A full-scale agent-based model to hypothetically explore the impact of lockdown, social distancing, and vaccination during the COVID-19 pandemic in Lombardy, Italy: Model development. JMIRx med 2(3):24630

Haghighat F (2021) Predicting the trend of indicators related to COVID-19 using the combined MLP-MC model. Chaos Solitons Fract 152:111399

Higgins JP, Li T, Deeks JJ (2019) Choosing effect measures and computing estimates of effect. Cochrane Handb Syst Rev Interv 2:143–176

Jordan D (2019) 1918 Pandemic (H1N1 virus). Centers of Disease Control and Prevention (CDC). Accessed 2021-06-20

Kolozsvári LR, Bérczes T, Hajdu A, Gesztelyi R, Tiba A, Varga I, Ala'a B, Szőllősi GJ, Harsányi S, Garbóczy S et al (2021) Predicting the epidemic curve of the Coronavirus (SARS-CoV-2) disease (COVID-19) using artificial intelligence: an application on the first and second waves. Inf Med Unlock 25:100691

Kou L, Wang X, Li Y, Guo X, Zhang H (2021) A multi-scale agent-based model of infectious disease transmission to assess the impact of vaccination and non-pharmaceutical interventions: The COVID-19 case. J Saf Sci Resilie 2(4):199–207

Kuo C-P, Fu JS (2021) Evaluating the impact of mobility on COVID-19 pandemic with machine learning hybrid predictions. Sci Total Environ 758:144151

Lasserson TJ, Thomas J, Higgins JP (2019) Starting a review. Cochrane Handb Syst Rev Interv 2:1–12

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. PLOS Med 6(7):1–28. https://doi.org/10.1371/journal.pmed.1000100

Majhi R, Thangeda R, Sugasi RP, Kumar N (2021) Analysis and prediction of COVID-19 trajectory: a machine learning approach. J Public Aff 21(4):2537

Malakar S (2021) Geospatial modelling of COVID-19 vulnerability using an integrated Fuzzy MCDM approach: a case study of West Bengal, India. Model Earth Syst Environ 23:1–14

Mallick P, Bhowmick S, Panja S (2022) Prediction of COVID-19 infected population for Indian States through a State Interaction Network-based SEIR Epidemic Model. Ifac-papersonline 55(1):691–696

Marzouk M, Elshaboury N, Abdel-Latif A, Azab S (2021) Deep learning model for forecasting COVID-19 outbreak in Egypt. Process Saf Environ Prot 153:363–375

McGuinness LA (2019) RobVis: An R package and web application for visualising risk-of-bias assessments. https://github.com/mcguinlu/robvis

Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLOS Med 6(7):1–6. https://doi.org/10.1371/journal.pmed.1000097

Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 4(1):1–9

Nama N, Sampson M, Barrowman N, Sandarage R, Menon K, Macartney G, Murto K, Vaccani J-P, Katz S, Zemek R et al (2019) Crowdsourcing the citation screening process for systematic reviews: validation study. J Med Internet Res 21(4):12953

Namasudra S, Dhamodharavadhani S, Rathipriya R (2021) Nonlinear neural network based forecasting model for predicting COVID-19 cases. Neural Process Lett 25:1–21

Nobi A, Tuhin KH, Lee JW (2021) Application of principal component analysis on temporal evolution of COVID-19. PLoS ONE 16(12):0260899

Noel-Storr A (2019) Working with a new kind of team: harnessing the wisdom of the crowd in trial identification. EFSA J 17:170715

Ohi AQ, Mridha M, Monowar MM, Hamid MA (2020) Exploring optimal control of epidemic spread using reinforcement learning. Sci Rep 10(1):1–19

Ozik J, Wozniak JM, Collier N, Macal CM, Binois M (2021) A population data-driven workflow for COVID-19 modeling and learning. Int J High Perform Comput Appl 35(5):483–499

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372 https://www.bmj.com/content/372/bmj.n71.full.pdf. https://doi.org/10.1136/bmj.n71

Pan Y, Zhang L, Yan Z, Lwin MO, Skibniewski MJ (2021) Discovering optimal strategies for mitigating COVID-19 spread using machine learning: experience from asia. Sustain Cities Soc 75:103254

Pang S, Hu X, Wen Z (2021) Environmental risk assessment and comprehensive index model of disaster loss for COVID-19 transmission. Environ Technol Innov 2:101597

Pourghasemi HR, Pouyan S, Heidari B, Farajzadeh Z, Shamsi SRF, Babaei S, Khosravi R, Etemadi M, Ghanbarian G, Farhadi A et al (2020) Spatial modeling, risk mapping, change detection, and outbreak trend analysis of Coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020). Int J Infect Dis 98:90–108

Raheja S, Kasturia S, Cheng X, Kumar M (2021) Machine learning-based diffusion model for prediction of Coronavirus-19 outbreak. Neural Comput Appl 2:1–20

Rashed EA, Hirata A (2021) Infectivity upsurge by COVID-19 viral variants in Japan: evidence from deep learning modeling. Int J Environ Res Public Health 18(15):7799

Sah S, Surendiran B, Dhanalakshmi R, Mohanty SN, Alenezi F, Polat K (2022) Forecasting COVID-19 pandemic using Prophet, ARIMA, and hybrid stacked LSTM-GRU models in India. Comput Math Methods Med 2022:2

Schunemann H (2013) GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group, 2013

Senthilkumar Mohan JA, Abugabah A, Adimoolam M, Singh SK, kashif Bashir A, Sanzogni L (2021) An approach to forecast impact of COVID-19 using supervised machine learning model. Software

Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, Li F (2020) Cell entry mechanisms of SARS-CoV-2. Proc Natl Acad Sci 117(21):11727–11734. https://doi.org/10.1073/pnas.2003138117

Shoaib M, Raja MAZ, Sabir MT, Bukhari AH, Alrabaiah H, Shah Z, Kumam P, Islam S (2021) A stochastic numerical analysis based on hybrid NAR-RBFs networks nonlinear SITR model for novel COVID-19 dynamics. Comput Methods Programs Biomed 202:105973

Shufro C (2020) How contact tracing breaks the chain of COVID-19 transmission. Hopkins Bloomberg Public Health . Accessed 2021-09-09

Swaraj A, Verma K, Kaur A, Singh G, Kumar A, de Sales LM (2021) Implementation of stacking based ARIMA model for prediction of COVID-19 cases in India. J Biomed Inform 121:103887

Syeda HB, Syed M, Sexton KW, Syed S, Begum S, Syed F, Prior F, Yu F Jr (2021) Role of machine learning techniques to tackle the COVID-19 crisis: systematic review. JMIR Med Inform 9(1):23811

Thomas J, Kneale D, McKenzie JE, Brennan SE, Bhaumik S (2019) Determining the scope of the review and the questions it will address. Cochrane Handb Syst Rev Interv 2:13–31

Ullah A, Wang T, Yao W (2022) Nonlinear modal regression for dependent data with application for predicting COVID-19. J R Stat Soc Ser A 185(3), 1424–1453. https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssa.12849. https://doi.org/10.1111/rssa.12849

Vasconcelos GL, Brum AA, Almeida FA, Mâcedo AM, Duarte-Filho GC, Ospina R (2021) Standard and anomalous second waves in the COVID-19 pandemic. medRxiv

Waffenschmidt S, Knelangen M, Sieben W, Bühn S, Pieper D (2019) Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. BMC Med Res Methodol 19(1):1–9

Wang Z, Nayfeh T, Tetzlaff J, O'Blenis P, Murad MH (2020) Error rates of human reviewers during abstract screening in systematic reviews. PLoS One 15(1):0227742

Zivkovic M, Bacanin N, Venkatachalam K, Nayyar A, Djordjevic A, Strumberger I, Al-Turjman F (2021) COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach. Sustain Cities Soc 66:102669