AJAS / SPAS

ISSN 2316-0861

**African Journal of Applied Statistics**
Vol. 11 (2), 2024, pages 1577 - 1612.
DOI: https://dx.doi.org/10.16929/ajas/2024.1577.283

ISSN 2316-0861

# Deep learning-based methods for predicting COVID-19: A critical review

**Fréjus Stéphane Tinhoun** [1,*], **Têlé Jonas Doumatè**[1,2] **and Romain Glèlè Kakaï**[1]

[1]Laboratoire de Biomathématiques et d'Estimations Forestières (LABEF), Faculté des Sciences Agronomiques, Université d'Abomey-Calavi, 04 BP 1525, Cotonou, Bénin
[2]Faculté des Sciences et Techniques (FAST), Université d'Abomey-Calavi, 01 BP 4521, Cotonou, Bénin

Received on July 18, 2025; Accepted on August 08, 2025; Published online on November 15, 2025



**Abstract.** The COVID-19 pandemic, caused by SARS-CoV-2 and first identified in Wuhan, China, disrupted human life worldwide. Numerous modeling studies have aimed to understand its dynamics. This systematic review analyzes and summarizes deep learning-based techniques used to predict COVID-19 cases. A research approach is outlined to perform a systematic literature review, focusing on formulating research questions, defining search criteria, and extracting relevant data. In total, 102 papers were considered after meeting the inclusion and exclusion criteria. Regarding scaling methods, two were reported, with normalization being the most widely used (39.22%). [$\cdots$] (see complete abstract in page 1578)

**Key words:** machine learning; infectious diseases; optimization; systematic Review.
**AMS 2010 Mathematics Subject Classification Objects:** 68T05; 62M45; 62P10.

---

(*) Fréjus Stéphane Tinhoun: frejus.tinhoun@uac.bj
Têlé Jonas Doumatè: jonas.doumate@uac.bj
Romain Glèlè Kakaï: glele.romain@gmail.com

**Full Abstract in English.** The COVID-19 pandemic, caused by SARS-CoV-2 and first identified in Wuhan, China, disrupted human life worldwide. Numerous modeling studies have aimed to understand its dynamics. This systematic review analyzes and summarizes deep learning-based techniques used to predict COVID-19 cases. A research approach is outlined to perform a systematic literature review, focusing on formulating research questions, defining search criteria, and extracting relevant data. In total, 102 papers were considered after meeting the inclusion and exclusion criteria. Regarding scaling methods, two were reported, with normalization being the most widely used (39.22%). Eight hyperparameter selection processes were reported, with grid search (6.87%), Bayesian optimization (2.94%), and manual search (2.94%) being the most commonly applied. Regarding model validation strategies, split validation (91.2%) was the most commonly used. For COVID-19 prediction, findings revealed that long short-term memory networks were the most extensively used (48.31%). Other frequently used models included hybrid models, gated recurrent units, convolutional neural networks, recurrent neural networks, compartmental models enhanced with deep learning, multi-layer perceptrons, and deep neural networks. Regarding what was predicted, confirmed cases were the most frequently predicted (29.95%), followed by deaths, new cases, recoveries, epidemiological parameters, active cases, intensive care unit admissions, and hospitalizations. However, several papers failed to report valuable and anticipated information derived from deep learning models. Scaling methods were not reported in 57 articles (55.88%), and hyperparameter selection processes were only noted in 19.61% of the papers. Additionally, the software used was reported in only 29.41% of the studies. To ensure the reliability of COVID-19 predictions, researchers must explicitly document the conditions under which deep learning models perform optimally and identify biases that could affect predictions.

**Résumé.** (Abstract in French) La pandémie de COVID-19, causée par le SARS-CoV-2 et identifiée pour la première fois à Wuhan, en Chine, a bouleversé la vie humaine dans le monde entier. De nombreuses études de modélisation ont cherché à comprendre sa dynamique. Cette revue systématique analyse et résume les techniques basées sur l'apprentissage profond utilisées pour prédire les cas de COVID-19. Une approche de recherche est décrite pour effectuer une revue systématique de la littérature, en mettant l'accent sur la formulation de questions de recherche, la définition de critères de recherche et l'extraction de données pertinentes. Au total, 102 articles ont été pris en compte après avoir satisfait aux critères d'inclusion et d'exclusion. En ce qui concerne les méthodes de mise à l'échelle, deux ont été signalées, la normalisation étant la plus largement utilisée (39,22%). Huit processus de sélection d'hyperparamètres ont été rapportés, la recherche par grille (6,87%), l'optimisation bayésienne (2,94%) et la recherche manuelle (2,94%) étant les plus couramment appliqués. En ce qui concerne les stratégies de validation des modèles, la validation par division (91,2%) était la plus couramment utilisée. Pour la prédiction de la COVID-19, les résultats ont révélé que les réseaux long short-term memory étaient les plus largement utilisés (48,31%). Parmi les autres modèles fréquemment utilisés figuraient les modèles hybrides, les unités récurrentes à portes, les réseaux neuronaux convolutifs, les réseaux neuronaux récurrents, les modèles compartimentaux améliorés par l'apprentissage profond, les perceptrons multicouches et les réseaux neuronaux profonds. En ce qui concerne les prédictions, les cas confirmés étaient les plus fréquemment prédits (29,95%), suivis des nouveaux cas, des décès, des guérisons, des paramètres épidémiologiques, des cas actifs, des admissions en soins intensifs et des hospitalisations. Cependant, plusieurs articles n'ont pas rapporté les informations précieuses et attendues issues des modèles d'apprentissage profond. Les méthodes de mise à l'échelle n'ont pas été mentionnées dans 57 articles (55,88%), et les processus de sélection des hyperparamètres n'ont été mentionnés que dans 19,51% des articles. De plus, le logiciel utilisé n'a été mentionné que dans 29,41% des études. Afin de garantir la fiabilité des prévisions relatives à la COVID-19, les chercheurs doivent documenter de manière explicite les conditions dans lesquelles les modèles d'apprentissage profond fonctionnent de manière optimale et identifier les biais susceptibles d'influencer les prévisions.

**Presentation of all authors**.

**Fréjus Stéphane TINHOUN**, M.sc, is preparing a Ph.D. degree at: Laboratoire de Biomathématiques et d'Estimations Forestières, Faculté des Sciences Agronomiques (FSA), Université d'Abomey-Calavi, 04, BP 1525, Cotonou, Bénin, under the supervision of the second and the third authors.

**Têlé Jonas DOUMATE**, Ph.D., is an Associate Professor in Applied Mathematics at Faculté des Sciences et Techniques (FAST), Université d'Abomey-Calavi, 01 BP 4521, Cotonou, Bénin.

**Romain GLELE KAKAÏ**, Ph.D., is a Full Professor of Biostatistics and Forest Estimations at Faculté des Sciences Agronomiques (FSA), Université d'Abomey-Calavi, 04 BP 1525, Cotonou, Bénin. He is a member of the African Academy of Sciences.

## 1. Introduction

The novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infectious disease, first identified in Wuhan, China, in December 2019, has spread rapidly worldwide (Raja *et al.*, 2021) with Africa's first confirmed case reported on February 14, 2020 (Organization, W.H., *et al.*, 2022, Haileamlak, 2021). It was declared a public health emergency by the World Health Organization (WHO) on January 20, 2020 (Jee, 2020) and a pandemic on March 11, 2020 (Organization, W.H., *et al.*, 2020). The pandemic has affected all aspects of life, most notably forcing people from both developing and developed countries to stay indoors (Reveil and Chen, 2022). In this context, information and communication technologies play a vital role in connecting communities, supporting policy implementation, and providing guidance through the analysis of the vast datasets generated during the COVID-19 pandemic (Arshad, 2021). SARS-CoV-2 is transmitted through direct contact with infected individuals or via contaminated objects and surfaces, similar to its predecessors, Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (Zhang *et al.*, 2020, Ahmed, 2015). To mitigate the spread of SARS-CoV-2 during the early stages of the pandemic, non-pharmaceutical interventions (NPIs) such as social distancing, contact tracing, quarantine, isolation, the use of face masks in public, city or nation-wide lockdowns, and travel restrictions were implemented (Niraula *et al.*, 2022, Perra, 2021). However, SARS-CoV-2 remains a concern despite the implementation of NPIs, with notable socio-economic consequences worldwide (Cho, 2020, Hevia and Neumeyer, 2020). Pharmaceutical interventions, including antivirals and vaccines, have been developed. Safe and effective vaccines, such as those from Pfizer, Moderna, Sputnik V, Sinovac, AstraZeneca, and Johnson & Johnson, were introduced by the end of 2020 (Iwry, 2021). However, novel SARS-CoV-2 variants have emerged and shown a partial ability to evade the protection offered by vaccines (Kustin *et al.*, 2021). Even with the availability of vaccines, SARS-CoV-2 continues to pose significant threats to global public health, with over 585 million reported cases and 6.4 million reported deaths as of August 2022 (Organization, W.H., *et al.*, 2022). In this chaotic situation, there is a pressing need for timely evaluation of the disease to ensure the successful coordination and mobilization of medical, human, and pharmaceutical resources (Surya, 2018). Around the world, researchers have studied the patterns of the COVID-19 pandemic, using conventional mathematical and statistical modeling techniques to model, predict, treat, and control the disease (Ngonghala *et al.*, 2020, Taboe *et al.*, 2020). However, these models have been limited in their accuracy for COVID-19 case predictions due to data quality issues (Ceylan, 2020, Jewell *et al.*, 2020). Deep learning (DL), a branch of machine learning (ML) within the field of artificial intelligence (AI), has gained prominence (Piccialli *et al.*, 2020, Xie *et al.*, 2020). DL methods can be categorized into supervised, semisupervised, and unsupervised learning. In supervised learning, models are trained on labeled data that include both inputs and corresponding outputs. In contrast, unsupervised learning involves training on unlabeled data, often resulting in more complex and less interpretable solutions. Semi-supervised learning combines both labeled and unlabeled data for training. DL is a modern tool that imitates human brain functions to analyze images, sounds, texts, and various types of data (Xie *et al.,* 2020). It has the ability to identify patterns and relationships in data, leading to less biased estimates (Piccialli *et al.,* 2020, Wiens and Shenoy, 2018). DL has been increasingly used in many fields, including public health and epidemiology (Gürsoy and Kaya, 2023). Unlike conventional statistical models, DL can capture complex, non-linear relationships in COVID-19 data, enabling adaptive learning and higher predictive accuracy in real-world scenarios. It is widely recognized that DL applications in the COVID-19 pandemic have yielded efficient results (Dogan *et al.,* 2021). Summarizing information from these studies is of great importance. While previous reviews have examined ML/AI techniques for the COVID-19 pandemic, most have focused on COVID-19 diagnosis from X-ray images and CT scans (Subramanian *et al.,* 2022, Arman *et al.,* 2022, Islam *et al.,* 2021, Sharma and Gupta, 2020). To date, and to the best of our knowledge, there has been no comprehensive study specifically focused on case data and DL-based methods to predict COVID-19 cases. Therefore, this review aims to fill this gap in the literature by conducting a systematic and critical review of DL-based studies published from January 1, 2020, to October 31, 2024, on COVID-19 prediction. The main objectives of this systematic review are to: (i) summarize trends in DL modeling techniques used to predict COVID-19 reproduction ratios, cases, severity, hospitalization, recoveries, and deaths; (ii) assess the reliability of predictions for COVID-19 cases, severity, hospitalizations, recoveries, and deaths; and (iii) identify research gaps in DL applications for predicting COVID-19 cases and propose future research directions. The rest of the paper is structured as follows: Section 2 provides an overview of DL techniques, Section 3 details the methodology, the main results are presented in Section 4 and a short conclusion is offered in Section 5 to finalize the paper.

## 2. Overview of deep learning techniques

DL is a subfield of ML within the domain of AI. It aims to simulate computer systems that mimic the human brain. DL relies on algorithms that learn multiple levels of representation to model complex relationships between data (Deng *et al.,* 2014).

### *2.1. Deep learning evolution*

The first generation of artificial neural networks (ANNs) was limited in terms of computational capabilities and consisted of perceptrons arranged in neural layers. This was followed by the development of backpropagation, which quantified the error rate and propagated the error backward through the network (Mathew *et al.,* 2014). To address the limitations of backpropagation, the restricted Boltzmann machine (RBM) was introduced, making learning easier. Over time, other networks were developed, continuing the evolution of DL models (Mathew *et al.,* 2014, Miikkulainen *et al.,* 2024). The evolution of DL models, alongside traditional models, and the performance comparison between traditional ML algorithms and DL algorithms are depicted in Fig. 1 and 2 (Mathew *et al.,* 2014), respectively. As the amount of data increases, the performance of DL models improves, while the performance of traditional ML algorithms stabilizes once the training data threshold is reached.

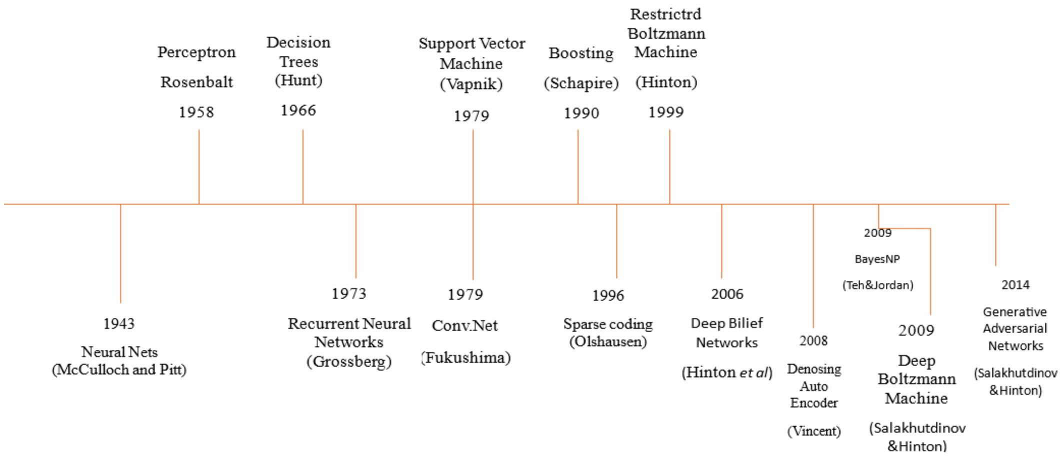


**Fig. 1.** Evolution of deep learning models.

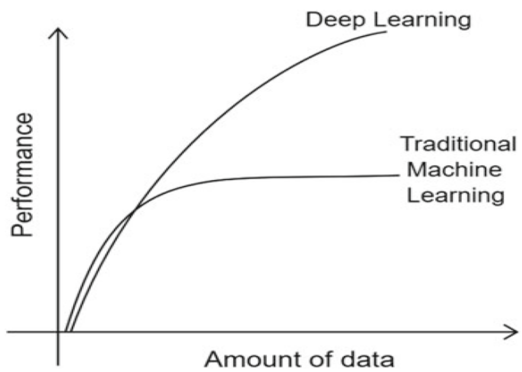


**Fig. 2.** Performance of Deep Learning Models and Traditional Machine Learning.

### *2.2. Deep learning approaches*

#### 2.2.1. Supervised Learning

In a supervised learning framework, the model is trained with labeled data, meaning that both input $(x_i)$ and output $(y_i)$ are supplied to the network. The input variables $(x_i)$ are mapped to output variables $(y_i)$ using an algorithm that learns the mapping function $f$ such that:

$$y_i = f(x_i). \quad (1)$$

The training error is used to improve the output.

#### 2.2.2. Unsupervised learning

In an unsupervised learning framework, the model is trained with unlabeled data, meaning that only the input is supplied to the network, and there is no output to map. This learning method seeks to understand the data by modeling their underlying distribution, which leads to the discovery of interesting structures within the data (Mathew *et al.*, 2014).

#### 2.2.3. Reinforcement learning

Reinforcement learning (RL) is a type of ML in which an agent learns to make decisions by performing actions in an environment to maximize cumulative rewards and minimize penalties. It was introduced in 2013 with Google DeepMind (Mnih *et al.*, 2015, Alzubaidi *et al.*, 2021). RL does not explicitly instruct the algorithm on how to learn but instead allows it to explore and solve problems independently (Mnih, 2016).

#### 2.2.4. Hybrid learning

Hybrid learning refers to a learning situation in which the model is trained using both labeled and unlabeled data by combining supervised and unsupervised architectures. This approach is expected to yield significantly better results for human action recognition, particularly when using action bank features (Deng *et al.*, 2014).

### *2.3. Architectures for deep learning*

DL architectures have outperformed basic ANNs. While deep structures are inefficient with respect to training time, this limitation can be addressed using transfer learning and graphics processing unit computing (Mathew *et al.*, 2014). The following are some relevant DL architectures:

#### 2.3.1. Unsupervised pre-trained networks

In this learning framework, the model is trained unsupervised and then used for prediction. The following unsupervised pre-training architectures are commonly used (Erhan *et al.*, 2010):

Autoencoders: In an autoencoder, the first layer is the encoder, and the transpose of the encoder is the decoder. During training, the input is reconstructed using an unsupervised method. After training, the model moves to subsequent layers of deep networks, fixing the weights of the previous layer. The network then returns to the original problem to be solved (e.g., classification or regression) and optimizes the model using stochastic gradient descent (SGD) starting with the weights obtained from pre-training. The encoder and decoder are the main components of an autoencoder network (Hubens, 2018). The input is translated by the encoder into a latent space representation, which can be described as follows:

$$h = f(x). \tag{2}$$

The decoder reconstructs the input from the latent space representation, which can be described as follows:

$$r = g(h). \tag{3}$$

Autoencoders, as described in equation (4), can be summarized as follows:

$$g[f(x)] = r. \tag{4}$$

Deep belief networks (DBNs): DBNs rely on their first layer to learn features, and the activation of subsequent layers continues until the final layer is reached. The RBM and feedforward networks are responsible for training and fine-tuning the layers of DBNs, respectively. Unlike other deep networks, DBNs learn hidden patterns globally, rather than learning complex patterns progressively through each layer (Salakhutdinov and Hinton, 2009).

Generative adversarial networks (GANs): Initiated by Ian Goodfellow, GANs consist of generator and discriminator networks. The discriminator validates the content generated by the generator. A GAN is regarded as a two-player minimax algorithm and uses convolutional neural networks (CNNs) and feedforward neural networks (Goodfellow *et al.*, 2014).

### 2.3.2. Convolutional neural networks

CNNs are primarily designed for image processing. They assign weights and biases to different objects in an image, allowing the network to distinguish one class from another. CNNs apply specific filters to capture the spatial and temporal relationships within an image (Miikkulainen *et al.*, 2024, Le *et al.*, 2015). Common CNN architectures include ResNet, ZFNet, GoogLeNet, AlexNet, VGGNet, and LeNet, among others.

### 2.3.3. Recurrent Neural Networks

Recurrent neural networks (RNNs), which are known for time series prediction, are an improved version of traditional neural networks. RNNs take the outcome of the previous time step (t-1) as input for the current time step (t) (Sherstinsky, 2020, Zhao *et al.*, 2019), using their hidden state. RNNs rely on their "memory" to store all information about previous computations (Sherstinsky, 2020, Smyl, 2020), making them useful for time series forecasting.

However, RNNs suffer from the vanishing gradient problem when learning long-term dependencies.

Let $(x_1, x_2, \cdots, x_T)$ represent a sequence of inputs, and $\varphi$ be a non-linear mapping function. The hidden state $(h_t)$ can be defined as follows:

$$\begin{cases} h_t = \varphi(h_{t-1}, x_t), \text{if } t \neq 0 \\ 0, \text{if } t = 0. \end{cases} \tag{5}$$

Updating $(h_t)$ at time t can be typically implemented as follows:

$$h_t = g(Wx_t + Uh_{t-1}), \tag{6}$$

where g is a smooth and bounded mapping, similar to a hyperbolic tangent (tanh) function, and W represents the parameter matrix.

Long short-term memory (LSTM) and gated recurrent units (GRU) are two advanced versions of RNNs designed to address the vanishing gradient problem associated with RNNs.

LSTM is an effective tool for tackling the vanishing gradient issue associated with traditional RNNs. It has the ability to add or remove information from the cell state through three multiplicative gates: the input gate $(i_t)$, the forget gate $(f_t)$, and the output gate $(o_t)$ (Sheikhi and Kowsari, 2023). The LSTM equations are as follows:

$$\begin{cases} f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{c_t} = tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \\ c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c_t} \\ o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \cdot tanh(c_t), \end{cases} \tag{7}$$

where $\sigma$ denotes the sigmoid function.

GRU is an improvement over LSTM with fewer parameters and operations (Sheikhi and Kowsari, 2023). It uses two gates—the update gate and the reset gate—rather than three gates, allowing it to overcome the vanishing gradient issue without the overhead of standard LSTMs (Sheikhi and Kowsari, 2023).

The *update gate* in a gated combines the functions of the *inpute gate* and *forget gate* from an LSTM, regulating the retention of past information and ensuring that important features are preserved without being lost or overwritten. This also creates shortcuts across time steps to mitigate the vanishing gradient problem (Sheikhi and Kowsari, 2023).

The mathematics behind GRU is expressed as follows:

$$\begin{cases} r_t = \sigma(w_r \cdot x_t + U_r \cdot h_{t-1} + b_r) \\ z_t = \sigma(w_z \cdot x_t + U_z \cdot h_{t-1} + b_z) \\ \tilde{h_t} = \phi_h(w_h \cdot x_t + U_h \cdot (r_t \odot h_{t-1}) + b_c) \\ h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h_t}, \end{cases} \tag{8}$$

where $x_t$, $h_t$ and $\tilde{h_t}$ are the input vector, output vector, and candidate activation vector, respectively. $z_t$, $r_t$ and $\odot$ represent the update gate vector, reset gate vector, and Hadamard product (element-wise multiplication), respectively. Finally, W, U, and b represent the parameter matrices (for both the input vector and the hidden state) and the bias vector, respectively.

## 3. Methodology

This research mainly focuses on the methodology of the Systematic and Critical Literature Reviews (SCLR). SCLR is a technique used to identify, evaluate, and interpret all existing works related to a particular research question, subject area, or field of interest (Brereton *et al.*, 2007). Consequently, this review identifies published papers that describe the use of DL approaches for predicting COVID-19 cases based on case data.

### *3.1. Search strategy*

COVID-19 outbreak modeling papers were searched from January 1, 2020, to October 31, 2024, using the following keywords: ”machine learning methods,” ”deep learning methods,” OR ”artificial intelligence methods” combined with terms such as ”coronavirus disease,” ”COVID-19 disease,” “corona disease,” “SARS virus disease,” “SARS-CoV-2 disease,” OR “2019-nCoV disease”. The search was conducted across the following scientific databases: Scopus, Web of Science, and Google Scholar. For additional relevant studies, the bibliographies of retrieved articles, as well as those of current reviews and texts, were also examined. This process identified 1,026 articles, including 302 duplicates. After duplicates were removed, the screening process was conducted by reviewing the title, abstract, and content of each paper to ensure relevance to the use of DL for predicting COVID-19 cases based on case data. Through this screening, 622 papers were excluded, including 422 that addressed topics unrelated to COVID-19 modeling and 200 that focused on risk factors, socio-economic factors, patient diagnosis models, human behavior in response to control measures, or were non-English written studies. The methodology used in this review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, as outlined by Gnanvi *et al.*, (2021), and is summarized in Fig. 3.

### *3.2. Defining Research Questions*

Research questions are essential for structuring systematic reviews. The primary focus of this review was to investigate the role of DL methods in predicting COVID-19 cases. Data, being a critical component of any scientific application, were categorized by type, size, and source. Additionally, we examined the DL techniques, scaling methods, hyperparameters, validation strategies, and metrics applied in the selected papers. Finally, we identified the difficulties encountered by researchers and practitioners. Table 1 reformulates each research question for clarity.

### *3.3. Literature synthesis and analysis*

The literature synthesis was conducted in three parts. The first part focuses on the characteristics of the selected papers. The second part involves a bibliometric analysis, which was performed using VOSviewer to analyze keyword co-occurrence networks (Van Eck and Waltman, 2011). VOSviewer provides various user-friendly visualizations, primarily for the analysis of bibliometric maps. The third part addresses the research questions directly. The frequency of each response was determined based on its occurrences, with similar terms combined to simplify the analysis. Results are presented using tables or bar charts. Additionally, R software was used for further data analysis via the 'ggplot' function in the 'ggplot2' package.

## **4. RESULTS**

### *4.1. Characteristics of selected articles*

Journal-Wise Categorization

In this research process, we aimed to include articles that adhered to the PRISMA guidelines, without considering the publisher. However, in terms of academic and applied research publications, all the databases used are well recognized and relevant. Fig. 4c shows the distribution of selected papers by publisher. Almost 38.24% of the selected papers were published by Elsevier, making it one of the most prominent platforms for publishing high-quality papers. This was followed by MDPI (14.71%), Springer (13.73%), PLS (4.90%), Wiley (4.90%), Frontiers Media SA (2.94%), and others (Taylor & Francis, Royal Society, BMC, IEEE), which made up 6.86% of the publications. However, 13.73% of the papers had unreported publisher information.

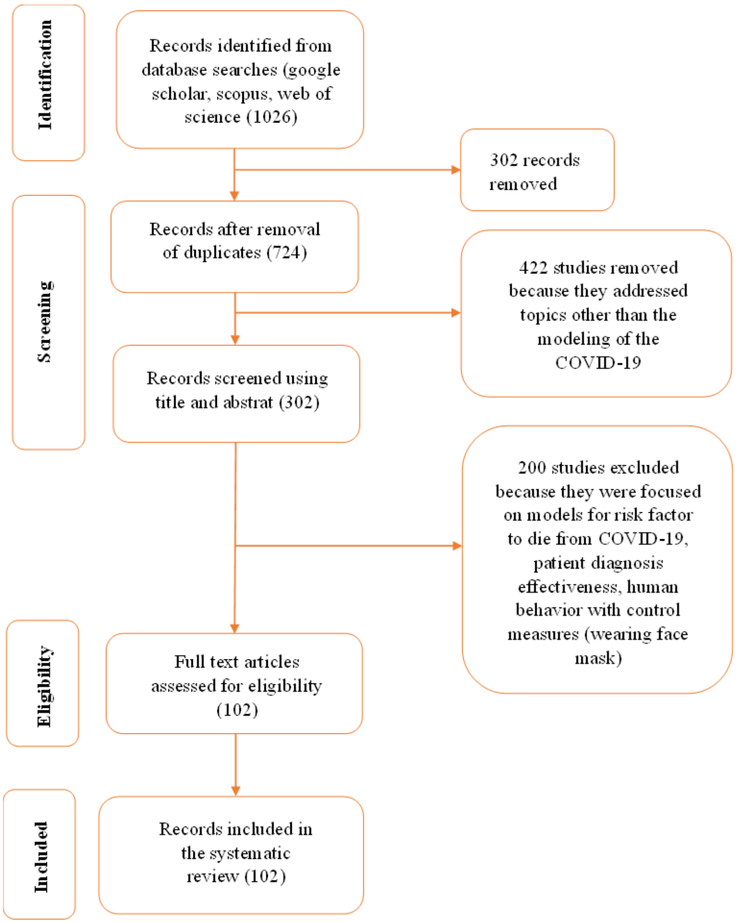


**Fig. 3.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the selection process.

*Continent-wise and country-wise statistics of data used in selected papers*

Fig. 4a and 4b show the distribution of data used across continents and the coverage of countries within each continent. The results show that Asia accounted for 71.13% of all studies, with India (35 studies) and China (18 studies) being the countries with the highest number of papers. This was followed by Europa (44.33%), the Americas (43.30%), Africa (18.56%), and Oceania (2.06%). Regarding country coverage within continents, Africa ranked highest, with studies conducted in 83.33% of its countries (45 out of 54). South Africa and Egypt had the most studies, with seven each. In the Europa, 31 out of 44 countries were represented in the selected papers. Additionally, some studies (Anno *et al.*, 2022, Miralles-Pechuán *et al.*, 2023, Zain and Alturki, 2021, Guo *et al.*, 2023) focused on global trends, providing a broader view of the pandemic.

**Table 1.** Research Questions Addressed in the Study

| **Number** | **Research Questions** |
|---|---|
| 1 | Why is deep learning used to help predict COVID-19 cases? |
| 2 | What type and source of data are in use? |
| 3 | What software platforms are in use? |
| 4 | To predict future COVID-19 cases, what deep learning techniques were proposed? |
| 5 | What scaling method is used? |
| 6 | What validation strategies were used? |
| 7 | What hyperparameters are used to implement the models? |
| 8 | What metrics were used to measure how the model performed? |
| 9 | What are the research gaps and future directions? |

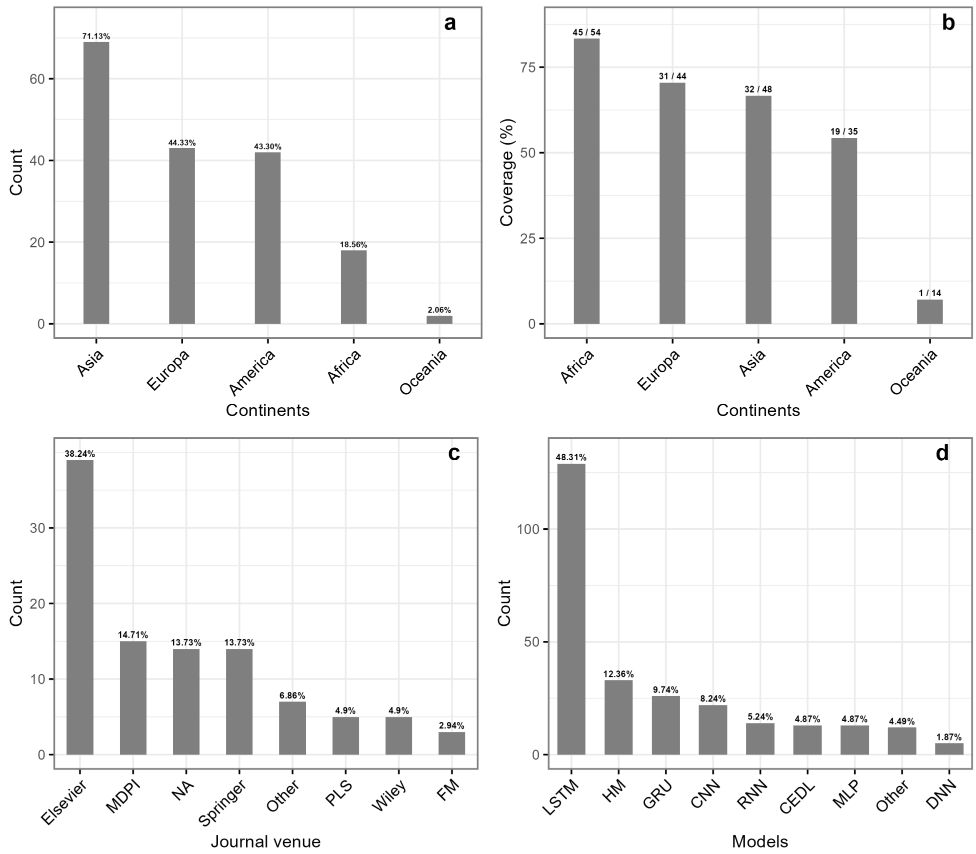


**Fig. 4.** Distribution of data used in selected articles across continents (a), country coverage across continents (b), distribution across publishers (c), and distribution of deep learning models (d).

### *4.2. Bibliometric analysis of keywords*

Table 2 presents the most frequent keywords that highlight the essential topics covered in the selected papers, along with their number of occurrences and total link strength. To improve accuracy, we resolved inconsistencies by consolidating similar keywords, such as ”LSTM” and ”long short-term memory”. The total link strength represents the overall strength of the co-occurrence links between a given keyword and other keywords. The keyword co-occurrence network is depicted in Fig. 6, which further clarifies the data presented in Table 2. Fig. 5 shows the trend in the publication years of the selected papers. The analysis reveals an increase in the use of DL techniques to predict COVID-19 cases from 2020 to 2021, followed by a decrease from 2021 to 2024.

Analyzing keywords associated with DL methods to predict COVID-19 cases reveals notable trends (Fig. 6). The widespread adoption of DL methods, with 44 occurrences, reinforces a data-driven strategy focused on predicting COVID-19 cases with less bias in disease management. LSTM gained significant prominence with 33 occurrences, highlighting its crucial role in COVID-19 disease prediction. Additionally, RNN and GRU were each used 6 times, indicating their usefulness in COVID-19 prediction. The application of CNN, primarily known for image processing, for predicting COVID-19 cases reveals the importance of its 1D version. Furthermore, the most important 17 keywords are grouped into four clusters by VOSviewer, as shown in Fig. 6.

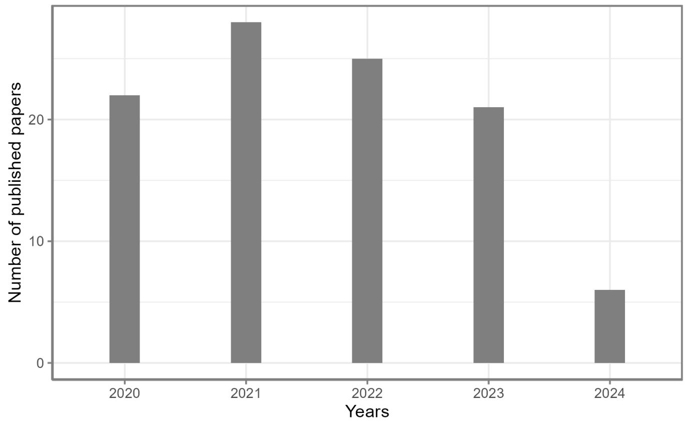


**Fig. 5.** Year-wise distribution of the articles

### *4.3. What is the reason of using DL in COVID-19 prediction?*

According to the authors, several reasons underlie the use of DL models for COVID-19 prediction. First, DL models are used to forecast the spread of infectious diseases and address the limitations of traditional methods. They have proven to be highly effective in providing better pattern analysis for both structured and unstructured data. Moreover, DL methods are used to predict COVID-19 because they can adapt to the nonlinear patterns in COVID-19 datasets and unrelated functions, potentially achieving state-of-the-art results with temporal data. They are also highly effective in implementing time series models. These models have outperformed traditional approaches by using cross-validation and improving performance through reduced uncertainty (Chimmula and Zhang, 2020). Additionally,

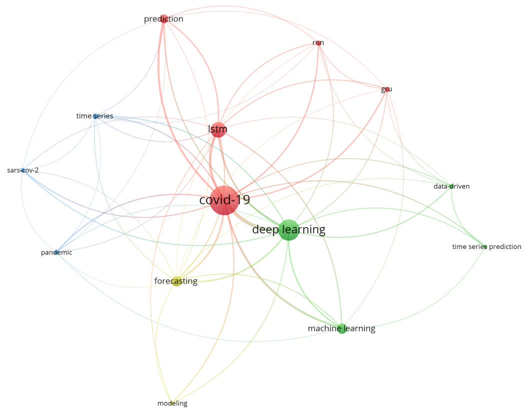


**Fig. 6.** Map of co-occurrence of the most used keywords network

**Table 2.** Most used keywords from the selected papers.

| **Keyword** | **Occurrences** | **Total link strength** |
|---|---|---|
| Covid-19 | 70 | 139 |
| Deep Learning | 44 | 102 |
| LSTM | 28 | 73 |
| Forcasting | 16 | 43 |
| Prediction | 13 | 35 |
| Machine learning | 16 | 31 |
| Time series | 7 | 24 |
| RNN | 6 | 21 |
| GRU | 6 | 20 |
| Pandemic | 5 | 15 |
| Data-driven | 5 | 14 |
| Sars-Cov-2 | 5 | 14 |
| CNN | 4 | 12 |
| Modeling | 4 | 10 |
| Time series prediction | 4 | 7 |
| Covid-19 pandemic | 4 | 4 |
| Covid-19 prediction | 4 | 4 |

DL models are used for public health security, supporting economic stability, and enabling the responsible allocation of medical resources.

### *4.4. What are the scaling method used?*

In time series data analysis, feature scaling of the input is an important step that significantly impacts the model's accuracy (Shahin and Almotairi, 2021, Al-Jabery *et al.*, 2019). Normalization and standardization are the two primary methods of feature scaling applied to time series data (Katris, 2021, Yan *et al.*, 2020).

Scaling features using normalization can be expressed mathematically as follows:

$$Y_n = \frac{O - O_{min}}{O_{max} - O_{min}}, \tag{9}$$

where $Y_n$ represents the normalized data, scaled between 0 and 1, for input data O. $O_{max}$ and $O_{min}$ represent the maximum and minimum values of the input O, respectively.

Scaling features using the standardization technique can be expressed mathematically as follows:

$$Y_s = \frac{O - \mu}{\sigma}, \tag{10}$$

where $Y_s$ represents the standardized data for the input data O. $\mu$ represents the mean value of the input O, and $\sigma$ represents the standard deviation of the input O.

Table 3 shows the scaling methods used in the reviewed papers. Normalization was the most commonly used scaling method, employed in 39.22%, of the selected papers, while 4.90% used feature standardization. The remaining, 55.88% of the papers did not report their method of feature scaling.

**Table 3.** Table of feature scaling methods used in DL models.

| Feature scaling methods | Number of article (%) |
|---|---|
| Normalization | 40 (39.22%) |
| Standardization | 5 (4.90%) |
| NR | 57 (55.88%) |

NR = Not Reported.

### *4.5. What are the hyperparameters optimization tuned method used?*

Optimization refers to the process of identifying the best model parameters during the training phase. Its goal is to minimize the loss function while maintaining the model's ability to generalize. However, the resulting model accuracy is sensitive to certain hyperparameters (e.g., number of epochs, batch size, and hidden layers) that are not optimized during training (Houetohossou *et al.*, 2023). Table 4 presents the optimization methods used in the selected papers. Among these, manual hyperparameter optimization becomes increasingly challenging as the search space expands. Of the assessed papers, 6.87% used the grid search algorithm, while both manual and Bayesian optimization algorithms were used in 2.94% of the papers. Particle swarm optimization and random search were used in 1.96% of the papers, and whale optimization, jellyfish search, and exhaustive search were used in 0.98%. The remaining 80.39% of the selected papers did not report the optimization algorithm used.

**Table 4.** Table of hyperparmeters optimization algorithm used.

| Optimization algorithms | Number of article (%) |
|---|---|
| Grid search | 7 (6.87%) |
| manual | 3 (2.94%) |
| bayesian optimization | 3 (2.94%) |
| particle swarm optimization | 2(1.96%) |
| Random search | 2 (1.96%) |
| Whale optimization | 1 (0.98%) |
| Jellyfish search | 1 (0.98%) |
| exhaustive search | 1 (0.98%) |
| NR | 82 (80.39%) |

NR = Not Reported.

### *4.6. What Validation Strategies were used?*

Fig. 7a presents the validation strategies used in the assessed papers and their respective proportions. The results reveal that the split validation strategy (91.2%) is the most extensively used, followed by the cross-validation strategy (8.8%). The widespread adoption of the split validation method is likely due to the limited availability of data.

### *4.7. What are the most frequently applied DL techniques and cases predict in COVID-19 modeling?*

All DL models used in the selected papers were trained within a supervised learning framework. Table 5 presents the percentage usage, strengths, and limitations of the top DL techniques used in the assessed papers.

Table 5: Percent usage, strengths and limitations of top models.

| Model | Percent usage | Strengths | Limitations |
|---|---|---|---|
| LSTM | 48.31% | Long-term dependency learner (Ahmed *et al.*, 2023). Insensitive to the vanishing gradient issue associated with RNNs (Ahmed *et al.*, 2023). Efficient for modeling time series data (Bedi *et al.*, 2021). Capable of extracting both long- and short-term correlations in time series, enabling the model to effectively identify data patterns (Greff *et al.*, 2016, Shrestha and Mahmood, 2019, Hoque and Aljamaan, 2021). | Computationally costly, increases training time and memory consumption due to its multiple gates and memory cells, leading to overfitting, especially with limited training data, and complicated to use due to its training instability, hyperparameter sensitivity, and parallelization constraints (Kandadi and Shankarlingam, 2025). |

| Model | Percent usage | Strengths | Limitations |
|---|---|---|---|
| HM | 12.36% | Better generalization performance of the final model, irrespective of unseen data (Ganaie *et al.*, 2022, Bozkurt, 2022). Ability to overcome challenges faced by conventional models (Ahmed *et al.*, 2023). Captures complex spatio-temporal patterns in the data and provides accurate forecasts (Tariq *et al.*, 2023, Khan *et al.*, 2022). Handles multi-dimensional data issues more effectively than traditional models (Wang *et al.*, 2018). Encodes spatiotemporal data in its memory cell (Wang *et al.*, 2018). Addresses big data challenges, including computational costs, high-dimensional sample sizes, storage, and error reduction (Azevedo *et al.*, 2024). Provides higher accuracy than traditional models with medium computational complexity (Ahmed *et al.*, 2023). Effectively manages unbalanced classes through data augmentation (Loukil, 2024). | Complexity arises from the combination of multiple learning techniques, with a high potential for overfitting due to the model's capacity. The model's complexity may demand significant computational resources, and it is heavily dependent on the accuracy of pre-processing and registration. There are also optimization challenges and potential computational overhead from hyperparameter tuning, particularly when integrating evolutionary computation. Model optimization and hyperparameter tuning are time-consuming, and the complexity in training and optimization may require substantial data and resources. Furthermore, it is not easy to interpret for clinical decision-making (Loukil, 2024). |

| Model | Percent usage | Strengths | Limitations |
|---|---|---|---|
| GRU | 9.74% | Suitable for modeling time dependencies and sequences (Dairi *et al.*, 2024). Shorter training time and fewer parameters than LSTM (Dairi *et al.*, 2024). | Slow convergence rate and low learning efficiency, leading to long training times and potential underfitting (Dairi *et al.*, 2024). |
| CNN | 8.24% | Automatic feature extraction from data (Xu *et al.*, 2022). Capable of capturing spatial relationships with strong generalization potential (Talaei Khoei *et al.*, 2023). | Requires a large amount of labeled data, with difficult parameter tuning and high computational costs (Talaei Khoei *et al.*, 2023). Critical tuning of hyperparameters is required (Loukil, 2024). |
| RNN | 5.24% | Sometimes achieves fast convergence with minimal parameters (Talaei Khoei *et al.*, 2023). Lower computational cost than LSTM (Zeroual *et al.*, 2020). | Vanishing and exploding gradient issues, along with difficult parameter tuning and poor spatial feature representations (Talaei Khoei *et al.*, 2023). |

Fig. 4d displays the DL models used in the assessed papers, which can be categorized into nine groups. The most frequently used predictive model was LSTM (48.31%). Other commonly used models included hybrid models (HMs), GRUs, CNNs, RNNs, compartmental models enhanced with deep learning (CEDL), multilayer perceptrons (MLPs), deep neural networks (DNNs), and other models, such as temporal convolutional networks, temporal fusion transformers, neural hierarchical interpolation for time series, transformer networks, and neural basis expansion analysis for time series. Recall that a compartmental model is one in which the population is divided into homogeneous sub-populations, such as susceptible, infectious, and recovered, representing different health states (Blackwood and Childs, 2018).

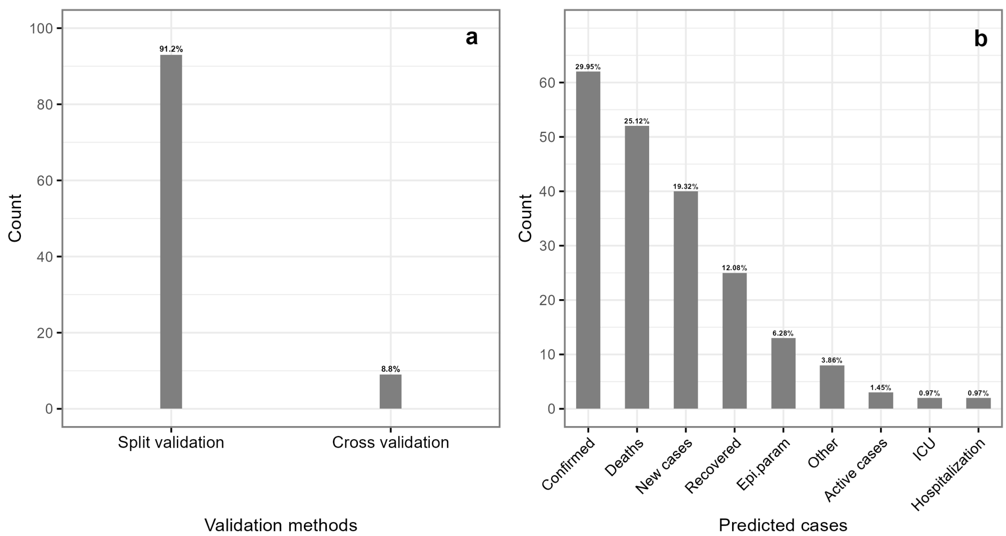


**Fig. 7.** Validation strategy (a) and COVID-19 case predicted (b).

Regarding prediction, diverse COVID19 cases were predicted and categorized into nine groups, as shown in Figure 7b. The most predicted case was confirmed cases (29.95%), followed by deaths (25.12%), new cases (19.32%), recovered cases (12.08%), epidemiological parameters (6.28%), active cases (1.45%), intensive care unit (ICU) cases (0.97%), hospitalization (0.97%), and other cases (3.86%), which include quarantine, peak date, epidemic final size, and susceptible–exposed–infected–recovered cases.

*4.8. What are the type and source of data used?*

The outcome of the modeling approach depends on the quality of the data. Therefore, data are considered a fundamental component in the application of scientific methods. Researchers generally adopt either open or closed sources as their primary data approaches (Dogan *et al.*, 2021, Shuja *et al.*, 2021). Closed-source data are regarded as exclusive to proprietary assets, while open-source data are associated with higher intrinsic value, transparency, verifiability, and usability (Xu *et al.*, 2020, Frazer *et al.*, 2020). Fig. 8 shows the distribution of data used in the selected papers, categorized by source, along with some statistics. Of the data used, 84.6% come from open sources while 15.4% are from closed sources (Fig. 8a). They were collected from distinct databases (Fig. 8b). The relative contribution of each source to the overall dataset is as follows: Government, Johns Hopkins University (JHU) Center for Systems Science and Engineering, World Health Organization (WHO), Our world in Data (OWD), Center for Disease Control and Prevention (CDC), Humanitarian Data Exchange (HDE), Kaggle and other (Real-world TSF, covid19india.org website, SNU ARIC) with 28.2%, 24.3%, 15.5%, 8.7%, 8.7%, 3.9%, 2.9%, 2.9% and 4.9%, respectively (Fig. 8b). Table 6 presents various features related to the use of data. All data used in the selected papers consist of case data, including new cases, confirmed cases, active cases, recoveries, deaths, quarantine data, mortality rates, and hospitalization, all within a defined timeframe.

**Table 6.** Table of sample size, training (rate and size), testing (rate and size), observation windows size and predictive windows size.
Abbreviations: **O.W.**: Observation windows; **P.W.**: Predictive windows

| | Sample size | Training | | Testing | | O.W. | P.W. |
|---|---|---|---|---|---|---|---|
| | | Rate (%) | Size | Rate (%) | Size | | |
| Min | 62 | 51 | 55 | 10 | 7 | 1 | 1 |
| Max | 1073 | 92 | 751 | 49 | 161 | 144 | 200 |

### *4.9. What software platforms are used?*

The software platforms used to run the DL models are presented in Table 7. As indicated by the results, 57.85% of the selected articles used Python, 8.82% used Matlab, 1.96% used R, and 1.96% used other software (such as UMPA and AMOR). The remaining 29.41% of the models in the selected articles did not report the software used.

### *4.10. What are the hyper-parameters used to implement the models?*

Table 8 presents relevant information regarding the DL hyperparameters in the assessed papers, in accordance with (Nabi *et al.*, 2021, Ayoobi *et al.*, 2021). These hyperparameters include the number of layers, hidden units, epochs, batch size, learning rate, optimizer, loss function, activation function, window size, kernel size and number of filters.

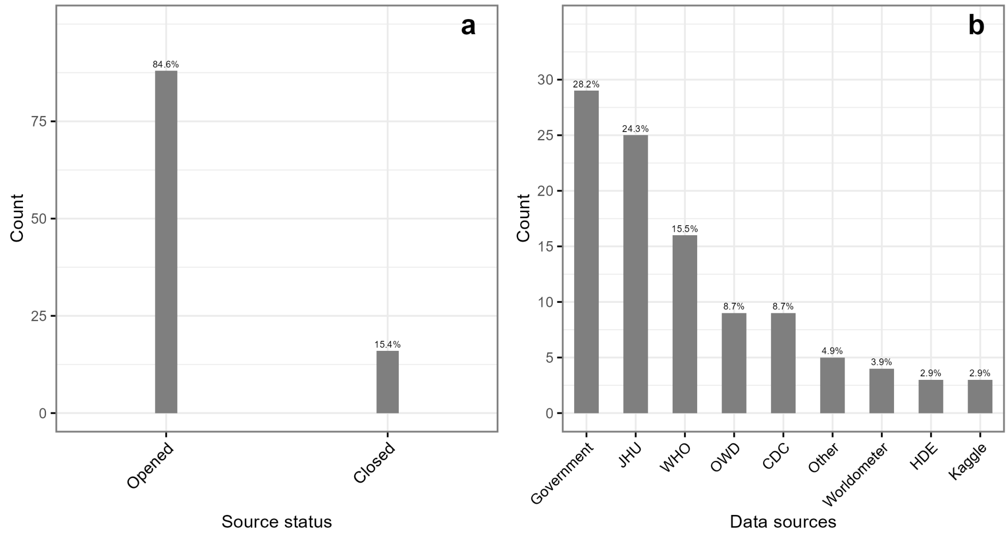


**Fig. 8.** Data source status and distribution of data.

**Table 7.** Table of software platforms used in DL models.

| Statistical software used | Number of articles (%) |
|---|---|
| Python | 59 (57.85%) |
| Matlab | 9 (8.82%) |
| R | 2 (1.96%) |
| Other | 2 (1.96%) |
| NA | 30 (29.41%) |

NA = Not Avaible.

Batch size refers to the number of training examples used in one iteration. Its impact on DL models has been well-established (Usmani *et al.*, 2023). Among the applied models, 49.44% explicitly reported the batch size used, while 50.56% did not (Table 8).

Learning is an essential component in any intelligent system, particularly during the training phase (Igiri *et al.*, 2023). In the DL framework, the learning rate is an essential hyperparameter that affects how the model's weights are updated during training. It is also recognized that a model's performance is sensitive to learning rate scheduling (Gonzalez, 2020). When a very large number of training samples is available, the model may struggle with generalization (Igiri *et al.*, 2023). The learning rate is a positive value, typically ranging from 0.0 to 1.0, and controls the step size at each iteration as the model moves toward minimizing the loss function. Additionally, it indicates the model's learning speed from the data. Of the models assessed, 65.54% reported the learning rate used, while the remaining 34.46% did not disclose the learning rate used to achieve the reported performance (Table 8).

The loss function is another crucial component of a DL model. It evaluates the model's predicted performance by measuring the difference between the predicted output and the actual target values. During training, the loss function is used to guide the optimization process by quantifying the error in the model's predictions. The efficiency of the optimization process depends on the choice of loss function. Given its importance in training, it is notable that 58.43% of the models in the assessed papers did not mention the loss function used, while 41.57% did report it (Table 8). Among those that mentioned the loss function, mean square error (MSE), mean absolute error (MAE), root mean square error (RMSE), cross entropy, relative error (RE) and Huber were used by 71.8%, 10.0%, 6.4%, 5.5%, 3.6% and 2.7% of the models, respectively (Fig. 9c).

**Table 8.** Table of hyperparameters status.

| **Hyperparameters** | **Report status** | **Number of models (%)** |
|---|---|---|
| Batch size | Yes | 132 (49.44%) |
| | No | 135 (50.56%) |
| Learning rate | Yes | 175 (65.54%) |
| | No | 92 (34.46%) |
| Loss function | Yes | 111 (41.57%) |
| | No | 156 (58.43%) |
| Optimizer | Yes | 165 (65.38%) |
| | No | 97 (36.33%) |
| Activation function | Yes | 139 (50%) |
| | No | 139 (50%) |
| Number of epoch | Yes | 165 (61.80%) |
| | No | 102 (38.20%) |
| Number of layers | Yes | 175 (65.54%) |
| | No | 92 (34.46%) |
| Number of hidden units | Yes | 157 (58.58%) |
| | No | 111 (41.42%) |
| Window size | Yes | 115 (43.07%) |
| | No | 152 (56.93%) |
| Kernel size | Yes | 30 (60%) |
| | No | 20 (40%) |
| Number of filters | Yes | 28 (56%) |
| | No | 22 (44%) |

In the DL framework, an optimizer is an algorithm used during the training phase to minimize the model's error or loss function. However, in 36.33% of the applied models, the optimizers were not reported (Table 8). Among those that were reported (65.38%), the adaptive moment estimation (Adam) optimizer was the most extensively used, with 79.41% (Fig. 9a). Additionally, there was low usage of other optimizers, such as root mean square propagation, SGD, Adamax (a variant of the Adam algorithm), Bayesian optimization (BO), AdaGrad, and other optimizers (ADLR and SGDM), with usage rates of 7.65%, 4.71%, 4.71%, 1.18%, 1.18%, and 1.18%, respectively (Fig. 9a).

The activation function is a mathematical function that determines the output of each neuron in a neural network. It transforms a linear relationship into a non-linear one, enabling the network to capture complex patterns in real-world data. Unfortunately, the activation functions used were not reported in 50% of the applied models (Table 8). Among those that were reported, the rectified linear unit (Relu) activation function was the most extensively used, with 55.40% (Fig. 9b). This was followed by the hyperbolic tangent (tanh), sigmoid, linear, and Gaussian error linear unit (gelu) activation functions, with usage rates of 19.4%, 13.7%, 10.8%, and 0.70%, respectively (Fig. 9b).

An epoch refers to the training situation in which all data samples have been exposed to the neural network for learning patterns. Once all samples have been processed, one epoch is considered complete. The performance of DL models typically correlates positively with an increase in the number of epochs (Ajayi and Ashi, 2023). The number of epochs used was reported in 61.80% of the applied models, while 38.20% did not report the number of epochs (Table 8).

In the DL context, a layer is a crucial building block that processes input data and passes them to subsequent layers, forming the network topology. Each layer often performs a specific type of transformation on its input data, which enables the model to learn complex patterns. The performance of DNNs is related to the number of trainable layers (Basha *et al.*, 2020). In this study, the number of layers was reported in 65.54% of the applied models, while the remaining 34.46% did not report the number of layers used (Table 8), despite its impact on model performance.

The number of hidden units is defined as the number of neurons in a hidden layer of a neural network. It is a key component of DL models, performing the necessary computational tasks that allow the network to learn from data and make predictions. Some previous studies have revealed the sensitivity of DL models' performance to the number of hidden units (Berglund *et al.*, 2015, Sun *et al.*, 2016). The number of hidden units was documented in 58.58% of the models used, while 41.42% did not report the number of hidden units (Table 8).

Window size, which significantly impacts the accuracy of DL models, refers to the number of time steps in a time series to be used as input for the prediction model (Azlan *et al.*, 2019). In the models used, the window size was not reported

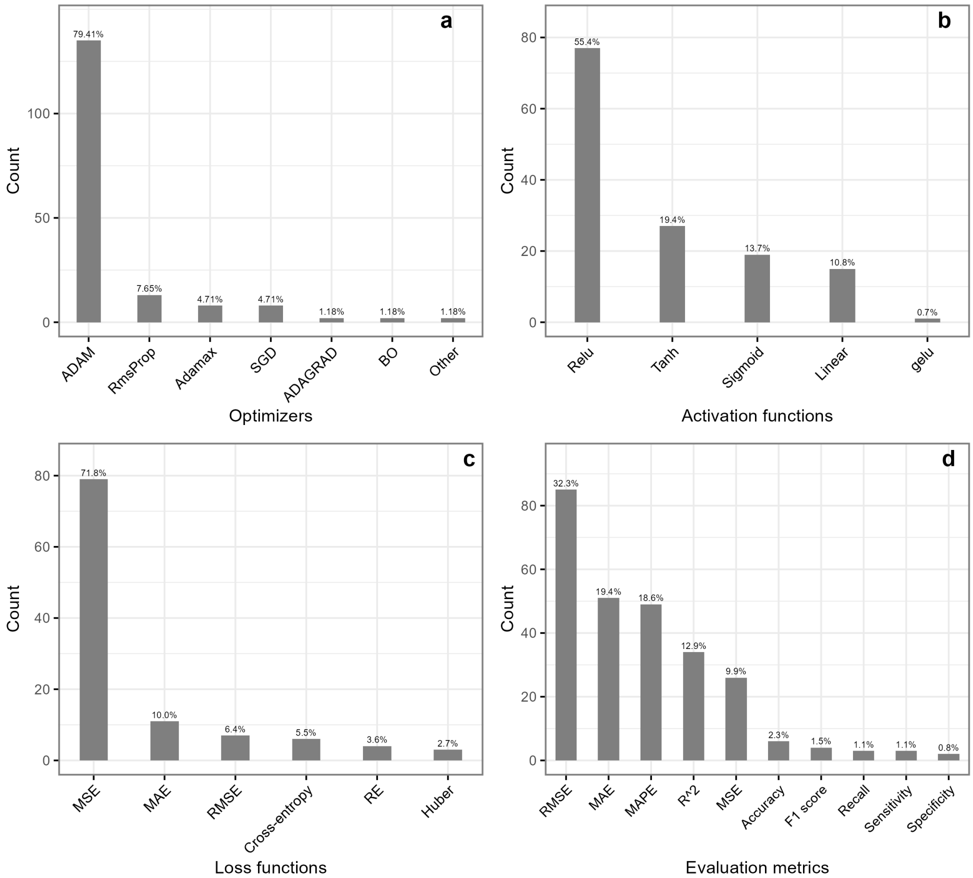


**Fig. 9.** List of optimizer used in selected papers (a), list of activation function used in selected papers (b), list of loss function used in selected papers (c) and list of evaluation metric used in selected papers (d).

in 56.93% of cases, compared to 43.07% where the window length was reported (Table 8).

In the context of CNNs, kernel size refers to the dimensions of the filter used to perform convolutional operations. The kernel size has an impact on the performance of CNNs, regardless of the architecture (Agrawal and Mittal, 2020, Chen and Shen, 2017). It was reported in 60% of the CNNs used in the selected papers, while in 40% of the models, the kernel size was not reported (Table 8).

In a CNN, filters are fundamental components used to detect patterns in the input data. During training, filters with smaller dimensions tend to result in less loss than those with larger dimensions (Khanday and Ashi, 2020). A total of 56% of the models used reported the number of filters, while 44% did not specify the number of filters (Table 8).

### *4.11. What metrics were used to measure how the model performed?*

The evaluation metrics enable scholars to measure and assess the quality of the work presented in any study, as well as effectively present results (Langfeldt *et al.,* 2021). Their choice is intrinsically linked to the applied models. Fig. 9d depicts the most commonly used regression metrics for model assessment in the selected papers. These metrics include root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), coefficient of determination ($R^2$), mean percentage error (MPE), mean square error (MSE), as well as recall, accuracy, sensitivity, $F_{score}$, and specificity (used less frequently), with usage rates of 32.3%, 19.4%, 18.6%, 12.9%, 9.9%, 2.3%, 1.5%, 1.1%, 1.1% and 0.8% (Fig. 9d). The aforementioned metrics are not necessarily the best evaluation metrics, but are mentioned to highlight the range of potential metrics that can be applied in DL-based methods for predicting COVID-19 based on case data.

RMSE represents the error between the predicted $\hat{f}(t)$ and observed $f(t)$ data and can be calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(f_t - \hat{f}_t)^2}. \tag{11}$$

$R^2$ represents the square of the correlation between the predicted and observed data and has values ranging from 0.0 to 1.0. It can be expressed as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(f_t - \hat{f}_t)^2}{\sum_{i=1}^{n}(\bar{f}_t - f_t)^2}. \tag{12}$$

MAE is used to assess the model's accuracy and goodness of fit for the datasets. It is the average absolute difference between the predicted value $\hat{f}(t)$ and the desired (correct) value $f(t)$. MAE can be calculated as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|f_t - \hat{f}_t|\cdot \tag{13}$$

MSE is the average squared difference between the predicted value $\hat{f}(t)$ and the desired (correct) value $f(t)$. It can be calculated as follows:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(f_t - \hat{f}_t)^2. \tag{14}$$

MAPE is the average percentage difference between the predicted and observed values, calculated as follows:

$$MAPE = 100 \times \frac{1}{n}\sum_{t=1}^{n}|\frac{f_t - \hat{f}_t}{f_t}|\cdot \tag{15}$$

The following evaluation metrics were used less frequently compared to the ones mentioned above:

$$Recall = \frac{tp}{tp + fn} \tag{16}$$

$$Precision = \frac{tp}{tp + fp} \tag{17}$$

$$Accuracy = \frac{tp + tn}{tp + tn + fn + fp} \tag{18}$$

$$F_{score} = 2 \times \frac{tp \times tn}{tp + fp}. \tag{19}$$

Despite their predictive power, DL models for COVID-19 prediction face challenges, such as data scarcity, overfitting to short-term trends, and difficulty generalizing across regions with varying outbreak dynamics. Future research should focus on exploring transfer learning techniques to address these issues.

### *4.12. What are the research gaps and future directions?*

Predicting COVID-19 cases during a pandemic is critically important and requires the use of advanced tools. This research highlights the significant progress made by scientists in developing DL algorithms to predict COVID-19 cases based on case data. However, limitations and gaps in research have hindered the creation of fully integrated COVID-19 case prediction systems. Firstly, we did not include articles published in the PubMed, Rxiv, and MedRxiv repositories due to their lack of peer review. Secondly, only English-written papers were included in this review. Third, articles were downloaded from Google Scholar, Web of Science, and Scopus databases for the review. However, the exclusion of published studies in non-English written (NEW) resulted in potentiel biais in outcomes related to systematic review studies (Jackson and Kuriyama, 2019). It is therefore important to include NEW articles in systematic review to ensure generalization and less biased outcomes. In spite of that, Walpole, (2019) pointed out that attention may be paid when removing language barrier, as it related to some challenges like costs, time and expertise of non-English languages. On the other hand, future studies could explore additional databases including non-peer ones as there is no significance improvement between preprints and peer-reviewed journal articles, specifically on COVID-19 trials (Davidson *et al.*, 2024). Additionally, Yadav and Shukla (2016) emphasized the importance and quality of the cross-validation strategy for model validation. However, this approach is rarely used due to the lack of large datasets. Therefore, in line with Saleem *et al.*, (2022), future research should focus on (a) making datasets available on public platforms, (b) utilizing data augmentation methods, and (c) evaluating the differences between the two validation strategies. Furthermore, because scaling methods, model hyperparameters, and hyperparameter optimization techniques affect the performance of DL models (Shahin and Almotairi, 2021, Al-Jabery *et al.*, 2019, Hoque and Aljamaan, 2021, Liao *et al.*, 2022), the results of future studies should specify the conditions under which the research was carried out.

## 5. Conclusion

The COVID-19 pandemic has had a profound global impact. This systematic and critical review examines 102 studies from three databases to address the research questions. The primary goal of this research is to organize and synthesize relevant literature, making it more accessible for scholars, practitioners, and academics to understand the existing methods, applications, and datasets. A key contribution of this study is the identification of DL methods used for disease prediction, as well as the scaling methods, hyperparameter optimization processes, validation strategies, and the types of data and datasets utilized. Among the most extensively used models for combating the pandemic were LSTM, HM, GRU, CNN, RNN, CEDL, MLP, and DNN. The most frequently used scaling method was normalization, while grid search, Bayesian optimization, manual search, particle swarm optimization, and random search were the most commonly used methods for hyperparameter tuning. The effectiveness of these methods varied widely across studies. The split method was the most commonly used, with various assessment metrics used, including RMSE, MAE, MAPE, $R^2$, MSE, accuracy, $F_{score}$, recall, sensitivity, and specificity. The predictions covered a wide range of outcomes, such as confirmed cases, deaths, new cases, recoveries, epidemiological parameters, active cases, ICU admissions, and hospitalizations. Future research should focus on systematically comparing hyperparameter tuning methods and validation strategies to improve the robustness of DL-based COVID-19 predictions.

*Acknowledgements*

This work was supported by the German Academic Exchange Service (DAAD). R Glèlè Kakaï acknowledges the Sub-Saharan Africa Advanced Consortium for Biostatistics (SSACAB) Phase II. We also thanks to the editorial board, the Associate Editor, and the anonymous referees for their supports and helpful comments.

*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.

## References

Raja, M.A., Mendoza, M.A., Villavicencio, A., Anjan, S., Reynolds, J.M., Kittipibul, V., Fernandez, A., Guerra, G., Camargo, J.F., Simkins, J. and others (2021). COVID-19 in solid organ transplant recipients: a systematic review and meta-analysis of current literature. *Transplantation Reviews*, 35, 100588.

Organization, W.H. and others (2022). Report on the strategic response to COVID-19 in the WHO African Region–1 February 2021 to 31 January 2022.

Haileamlak A. (2021). The impact of COVID-19 on health and health systems. SPAS-EDS. Saint-Louis of Senegal. *Ethiopian Journal of Health Sciences*, 31, 1073.

Jee Y. (2020). WHO international health regulations emergency committee for the COVID-19 outbreak. SPAS-EDS. Saint-Louis of Senegal. *Epidemiology and health*, 42.

Organization, W.H. and others, 2020. *WHO Director-General's opening remarks at the media briefing on COVID-19-11*, March 2020.

Reveil, M., Chen, Y-H. (2022). Predicting and preventing COVID-19 outbreaks in indoor environments: An agent-based modeling study. *Scientific Reports*, 12(1), 16076.

Arshad, M., (2022). The Role of Information and Communication Technology to Combat COVID-19 Pandemic: Emerging Technologies, Recent Developments and Open Challenges. *International Journal of Computer Science & Network Security*, 21(4), 93–102.

Langfeldt, L., Reymert, I., Aksnes, D.W., (2021). The role of metrics in peer assessments. *Research Evaluation*, 30(1), 112–126.

Saleem, F., Al-Ghamdi, A.S.A., Alassafi, M.O., AlGhamdi, S.A. (2022). Machine learning, deep learning, and mathematical models to analyze forecasting and epidemiology of COVID-19: a systematic literature review. *Transplantation Reviews*, 19, 5099.

Zhang, S.X., Liu, J., Jahanshahi, A.A., Nawaser, K., Yousefi, A., Li, J., Sun, S. (2020). At the height of the storm: Healthcare staff's health conditions and job satisfaction and their associated predictors during the epidemic peak of COVID-19. *Brain, behavior, and immunity*, 87, 144–146.

Ahmed M. (2015). Non-alcoholic fatty liver disease in 2015. *World journal of hepatology*, 7(11), 1450.

Niraula, P., Mateu, J., Chaudhuri, S. (2022). A Bayesian machine learning approach for spatio-temporal prediction of COVID-19 cases. *Stochastic Environmental Research and Risk Assessment*, 36(8), 2265–2283.

Perra N. (2021). Non-pharmaceutical interventions during the COVID-19 pandemic: A review. *Physics Reports*, 913, 1–52.

Cho S-W. (2020). Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden. *The Econometrics Journal*, 23(3). 323–344.

Hevia C. and Neumeyer A. (2020). A conceptual framework for analyzing the economic impact of COVID-19 and its policy implications. *UNDP Lac COVID-19 Policy Documents Series*, 1, 29.

Iwry J.L. (2021). FDA Emergency Use Authorization From 9/11 to COVID-19: Historical Lessons and Ethical Challenges. *Food & Drug LJ*, 76, 337.

Kustin, T., Harel, N., Finkel, U., Perchik, S., Harari, S., Tahor, M., Caspi, I., Levy, R., Leshchinsky, M., Ken Dror, S., and Others (2021). Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2-mRNA-vaccinated individuals. *Nature medicine*, 27(8), 1379–1384.

Surya L. (2018). How government can use AI and ML to identify spreading infectious diseases. *International Journal of Creative Research Thoughts (IJCRT), ISSN*, 2320–2882.

Ngonghala, C.N., Iboi, E., Eikenberry, S., Scotch, M., MacIntyre, C.R., Bonds, M.H., Gumel, A.B. (2020). Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel Coronavirus. *Mathematical biosciences*, 325, 108364.

Taboe, H.B., Salako, K.V, Tison, J.M., Ngonghala, C.N., Kakaï, R.G (2020). Predicting COVID-19 spread in the face of control measures in West Africa. *Mathematical biosciences*, 328, 108431.

Ceylan Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of The Total Environment*, 729, 138817.

Jewell, N.P.,Lewnard, J.A., Jewell, B.L. (2020). Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. *Jama*, 323(19), 1893–1894.

Piccialli, F., Di Somma, V., Giampaolo, F., Cuomo, S., Fortino, G. (2020). A survey on deep learning in medicine: Why, how and when?. *Information Fusion*, 66, 111–137.

Xie, X., Muruato, A., Lokugamage, K.G., Narayanan, K., Zhang, X., Zou, J., Liu, J., Schindewolf, C., Bopp, N.E., Aguilar, P.V., and Others (2020). An infectious cDNA clone of SARS-CoV-2. *Nature medicine*, 27(5), 841–848.

Wiens J. and Shenoy E.A. (2018). Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. *Clinical Infectious Diseases*, 66(1), 149–153.

Gürsoy, E. and Kaya, Y. (2023). An overview of deep learning techniques for COVID-19 detection: methods, challenges, and future works. *Multimedia Systems*, 29(3), 1603–1627.

Dogan, O., Tiwari, S., Jabbar, M.A., Guggari, S. (2021). A systematic review on AI/ML approaches against COVID-19 outbreak. *Complex & Intelligent Systems*, 7, 2655–2678.

Subramanian, N., Elharrouss, O., Al-Maadeed, S., Chowdhury, M. (2022). A review of deep learning-based detection methods for COVID-19. *Computers in Biology and Medicine*, 143, 105233.

Arman, S.E., Rahman, S., Deowan, S.A. (2022). A Bayesian optimization-based deep learning approach to diagnose COVID-19 from X-Ray images. *SN Computer Science*, 3(2), 115.

Islam, M.M., Karray, F., Alhajj, R., Zeng, J. (2021). A Bayesian optimization-based deep learning approach to diagnose COVID-19 from X-Ray images. *Ieee Access*, 9, 30551–30572.

Sharma S. and Gupta Y.K. (2020). Predictive analysis and survey of COVID-19 using machine learning and big data. *Journal of Interdisciplinary Mathematics*, (24)1, 175–195.

Deng, L., Yu, D., and others (2014). Deep learning: methods and applications. *Foundations and trends® in signal processing*, 7(3–4), 197–387.

Mathew, A., Amudha, P., Sivakumari, S. (2014). Deep learning techniques: an overview. *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*, 599–608.

Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., Raju, B., Shahrzad, H., Navruzyan, A., Duffy, N., and Others (2024). Evolving deep neural networks. *Artificial intelligence in the age of neural networks and brain computing*, 269–287.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., and Others (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8, 1–74.

Mnih V. (2016). Asynchronous Methods for Deep Reinforcement Learning. *arXiv preprint arXiv:1602.01783*.

Erhan, D., Courville, A., Bengio, Y., Vincent, P. (2010). Why does unsupervised pre-training help deep learning?. *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 201–208.

Hubens N. (2016). Deep inside: Autoencoders-towards data science. *Google Scholar*.

Salakhutdinov R. and Hinton G. (2009). Semantic hashing. *International Journal of Approximate Reasoning*, (50)7, 969–978.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

Le, Q.V., and others (2015). A tutorial on deep learning part 2: Autoencoders, convolutional neural networks and recurrent neural networks. *Google Brain*, 20, 1–20.

Sherstinsky A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404, 132306.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R.X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 115, 213–237.

Smyl S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International journal of forecasting*, 36(1), 75–85.

Sheikhi F., and Kowsari Z. (2023). Time series forecasting of COVID-19 infections and deaths in Alpha and Delta variants using LSTM networks. *Plos one*, 18(10), e0282624.

Brereton, P., Kitchenham, B.A, Budgen, D., Turner, M., Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of systems and software*, 80(4), 571–583.

Gnanvi, J.E, Salako, K.V., Kotanmi, G.B., Kakaï, R.G. (2021). On the reliability of predictions on Covid-19 dynamics: A systematic and critical review of modelling techniques. *Infectious Disease Modelling*, 6, 258–272.

Van Eck N.J., and Waltman L. (2011). VOSviewer manual. *Manual for VOSviewer version*, 1(0).

Anno, S., Hirakawa, T., Sugita, S., Yasumoto, S. (2022). A graph convolutional network for predicting COVID-19 dynamics in 190 regions/countries. *Frontiers in public health*, 10, 911336.

Miralles-Pechuán L., and Kumar A., and Suárez-Cetrulo, A.L. (2023). Forecasting COVID-19 cases using dynamic time warping and incremental machine learning methods. *Expert Systems*, 40(6), e13237.

Zain, Z.M., and Alturki, N.M. (2021). COVID-19 pandemic forecasting using CNN-LSTM: a hybrid approach. *Journal of Control Science and Engineering*, 2021(1), 8785636.

Guo, Z., Lin, Q., Meng, X. (2023). A Comparative Study on Deep Learning Models for COVID-19 Forecast. *Healthcare*, 11(17), 2400.

Chimmula, V.K.R., and Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, solitons & fractals*, 135, 109864.

Shahin, A.I., and Almotairi, S. (2021). A deep learning BiLSTM encoding-decoding model for COVID-19 pandemic spread forecasting. *Fractal and Fractional*, 5(4), 175.

Al-Jabery, Z., Obafemi-Ajayi, T., Olbricht, G., Wunsch, D. (2019). Computational learning approaches to data analytics in biomedical applications.

Katris C. (2021). A time series-based statistical approach for outbreak spread forecasting: Application of COVID-19 in Greece. *Expert systems with applications*, 166, 114077.

Yan, B., Tang, X., Liu, B., Wang, J., Zhou, Y., Zheng, G., Zou, Q., Lu, Y., and Tu, W. (2020). An improved method of COVID-19 case fitting and prediction based on LSTM.

Houetohossou, S.C.A., Houndji, V.R., Hounmenou, C.G., Sikirou, R., Kakaï, R.L.G. (2023). Deep learning methods for biotic and abiotic stresses detection and classification in fruits and vegetables: State of the art and perspectives. *Artificial Intelligence in Agriculture*.

Ahmed, S.F., Alam, Md.S.B., Hassan, M., Rozbu, M.R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, ABM., and Gandomi, A.H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, 56(11), 13521–13617.

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 computer science*, 2, 1–24.

Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10),2222–2232.

Shrestha, A., and Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE access*, 7, 53040–53065.

Hoque, K.E., and Aljamaan, H. (2021). Impact of hyperparameter tuning on machine learning models in stock price forecasting. *IEEE Access*, 9, 163815–163830.

Kandadi, T., and Shankarlingam, G. (2025). DRAWBACKS OF LSTM ALGORITHM: A CASE STUDY. *Available at SSRN 5080605*.

Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M., Suganthan, P.N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115, 105151.

Bozkurt, F. (2022). A comparative study on classifying human activities using classical machine and deep learning methods. *Arabian Journal for Science and Engineering*, 47(2), 1507–1521.

Tariq, M.U., Ismail, S.B., Babar, M., Ahmad, A. (2023). Harnessing the power of AI: Advanced deep learning models optimization for accurate SARS-CoV-2 forecasting. *PloS one*, 18(7), e0287755.

Khan, I.U., Afzal, S., Lee, J.W. (2022). Human activity recognition via hybrid deep learning based model. *Sensors*, 22(1), 323.

Wang, L., Xu, X., Dong, H., Gui, R., Yang, R., Pu, F. (2018). Exploring convolutional LSTM for PolSAR image classification. *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 8452–8455.

Azevedo, B.F., Rocha, A.M.AC., Pereira, A.I. (2024). Hybrid approaches to optimization and machine learning methods: a systematic literature review. *Machine Learning*, 113(7), 4055–4097.

Loukil, Z. (2024). A Hybrid Approach to Intelligent Prediction of Medical Conditions A Framework for Advancing Medical Diagnostics through Novel Hybrid Deep Learning Models DenCeption and HyBoost for Enhanced Feature Extraction and Predictive Accuracy in Medical Image Analysis.

Dairi, A., Harrou, F., Zeroual, A., Hittawe, M.M., Sun, Y. (2021). Comparative study of machine learning methods for COVID-19 transmission forecasting. *Journal of biomedical informatics*, 118, 103791.

Xu, L., Magar, R., Farimani, A.B., Hittawe, M.M., Sun, Y. (2022). Forecasting COVID-19 new cases using deep learning methods. *Computers in biology and medicine*, 144, 105342.

Talaei Khoei, T., Ould Slimane, H., Kaabouch, N. (2023). Deep learning: systematic review, models, challenges, and research directions. *Neural Computing and Applications*, 35(31), 23103–23124.

Blackwood, J.C., and Childs, L.M. (2018). An introduction to compartmental modeling for the budding infectious disease modeler. *Letters in Biomathematics*, 5(1), 195–221.

Zeroual, A., Harrou, F., Dairi, A., Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. *Chaos, solitons & fractals*, 140, 110121.

Shuja, J., Alanazi, E., Alasmary, W., Alashaikh, A. (2021). COVID-19 open source data sets: a comprehensive survey. *Applied Intelligence*, 51(3), 1296–1325.

Xu, B., Kraemer, M.UG., Group, D.C. (2020). Open access epidemiological data from the COVID-19 outbreak. *the lancet. Infectious Diseases*, 20(5), 534.

Frazer, J.S., Shard, A., Herdman, J. (2020). Involvement of the open-source community in combating the worldwide COVID-19 pandemic: a review. *Journal of medical engineering & technology*, 44(4), 169–176.

Nabi, K.N., Tahmid, Md.T., Rafi, A., Kader, M.E., Haider, Md.A. (2021). Forecasting COVID-19 cases: A comparative analysis between recurrent and convolutional neural networks. *Results in Physics*, 24, 104137.

Ayoobi, N., Sharifrazi, D., Alizadehsani, R., Shoeibi, A., Gorriz, J.M., Moosaei, H., Khosravi, A., Nahavandi, S., Chofreh, A.G., Goni, F.A., and others (2021). Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods. *Results in Physics*, 27, 104495.

Usmani, I.A, Qadri, M.T., Zia, R., Alrayes, F.S., Saidani, O., Dashtipour, K. (2023). Interactive effect of learning rate and batch size to implement transfer learning for brain tumor classification. *Electronics*, 12(4), 964.

Igiri, C.P, Anyama, O.U., Silas, A.I., Alrayes, F.S., Saidani, O., Dashtipour, K. (2023). Effect of learning rate on artificial neural network in machine learning.

Gonzalez, S. (2020). Improving deep learning through loss-function evolution.

Blanc, G., Gupta, N., Valiant, G., Valiant, P. (2020). Implicit regularization for deep neural networks driven by an ornstein-uhlenbeck like process. *Conference on learning theory*, 483–513.

Ajayi, O.G., and Ashi, J. (2023). Effect of varying training epochs of a faster region-based convolutional neural network on the accuracy of an automatic weed classification scheme. *Smart Agricultural Technology*, 3, 100128.

Basha, SH.S., Dubey, S.R., Pulabaigari, V., Mukherjee, S. (2020). Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*, 378, 112–119.

Berglund, M., Raiko, T., Cho, K. (2015). Measuring the usefulness of hidden units in Boltzmann machines with mutual information. *Neural Networks*, 64, 12–18.

Sun, S., Chen, W., Wang, L., Liu, X., Liu, T-Y. (2016). On the depth of deep neural networks: A theoretical view. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1).

Azlan, A., Yusof, Y., Mohsin, M.F.M., Gupta, N., Jindal, V. (2019). Determining the impact of window length on time series forecasting using deep learning. *International Journal of Advanced Computer Research*, 9(44), 260–267.

Agrawal A., and Mittal N. (2020). Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. *The Visual Computer*, 36(2), 405–412.

Chen J., and Shen Y. (2017). The effect of kernel size of CNNs for lung nodule classification. *IEEE*, 340–344.

Khanday, O., and Dadvandipour, S. (2020). Convolutional neural networks and impact of filter sizes on image classification. *Multidiszciplináris Tudományok*, 10(1), 55–60.

Yadav, S., and Shukla, S. (2016). Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. *2016 IEEE 6th International conference on advanced computing (IACC)*, 78–83.

Liao, L., Li, H.,Shang, W., Ma, L., Liu, T-Y. (2022). An empirical study of the impact of hyper-parameter tuning and model optimization on the performance properties of deep neural networks. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 31(3), 1–40.

Jackson J. F., and Kourima A. (2019). How often do systematic reviews exclude articles not published in English? *Journal of general internal medicine*, 34(8), 1388–1389.

Davidson, M., Evrenoglou, T., Gra $\tilde{n}$ a, C., Chaimani, A., Boutron, I. (2024). Comparison of effect estimates between preprints and peer-reviewed journal articles of COVID-19 trials. *BMC medical research methodology*, 24(1), 9.

Walpole S.C. (2019). Including papers in languages other than English in systematic reviews: important, feasible, yet often omitted. *Journal of Clinical Epidemiology*, 111, 127–134.