REVIEW

Forecasting COVID-19 Pandemic – A scientometric Review of Methodologies Based on Mathematics, Statistics, and Machine Learning

Pronóstico de la pandemia de COVID-19: una revisión cienciométrica de metodologías basadas en matemáticas, estadística y aprendizaje automático

Satya Prakash[1] ✉, Anand Singh Jalal[2], Pooja Pathak[1]

[1]GLA University, Department of Computer Engineering and Applications. Mathura, UP, INDIA, 281406.
[2]Devi Ahilya Vishwavidyalaya (DAVV), School of Computer Science & Information Technology. Indore, MP, INDIA, 452001.

ABSTRACT

Introduction: the COVID-19 pandemic is being regarded as a worldwide public health issue. The virus has disseminated to 228 nations, resulting in a staggering 772 million global infections and a significant death toll of 6,9 million. Since its initial occurrence in late 2019, many approaches have been employed to anticipate and project the future spread of COVID-19. This study provides a concentrated examination and concise evaluation of the forecasting methods utilised for predicting COVID-19.
Method: to begin with, A comprehensive scientometric analysis has been conducted using COVID-19 data obtained from the Scopus and Web of Science databases, utilising bibliometric research. Subsequently, a thorough examination and classification of the existing literature and utilised approaches has been conducted. First of its kind, this review paper analyses all kinds of methodologies used for COVID-19 forecasting including Mathematical, Statistical, Artificial Intelligence - Machine Learning, Ensembles, Transfer Learning and hybrid methods.
Results: data has been collected regarding different COVID-19 characteristics that are being taken into account for prediction purposes, as well as the methodology used to develop the model. Additional statistical analysis has been conducted using existing literature to determine the patterns of COVID-19 forecasting in relation to the prevalence of methodologies, programming languages, and data sources.
Conclusions: this review study may be valuable for researchers, specialists, and decision-makers concerned in administration of the Corona Virus pandemic. It can assist in developing enhanced forecasting models and strategies for pandemic management.

Keywords: Forecasting; COVID-19; Time Series Prediction; Machine Learning; Mathematical Models; Data Driven Models.

RESUMEN

Introducción: la pandemia de COVID-19 está siendo considerada un problema de salud pública a nivel mundial. El virus se ha diseminado a 228 países, lo que ha provocado la asombrosa cifra de 772 millones de infecciones globales y un importante número de muertes de 6,9 millones. Desde su aparición inicial a finales de 2019, se han empleado muchos enfoques para anticipar y proyectar la futura propagación de la COVID-19. Este estudio proporciona un examen concentrado y una evaluación concisa de los métodos de pronóstico utilizados para predecir COVID-19.
Método: para empezar, se realizó un análisis cienciométrico integral utilizando datos de COVID-19 obtenidos de las bases de datos Scopus y Web of Science, utilizando investigación bibliométrica. Posteriormente, se

llevó a cabo un examen y una clasificación exhaustivos de la literatura existente y los enfoques utilizados. Este artículo de revisión, el primero de su tipo, analiza todo tipo de metodologías utilizadas para el pronóstico de COVID-19, incluidas matemáticas, estadísticas, inteligencia artificial: aprendizaje automático, conjuntos, aprendizaje por transferencia y métodos híbridos.

Resultados: se han recogido datos sobre las diferentes características de la COVID-19 que se están teniendo en cuenta a efectos de predicción, así como la metodología utilizada para desarrollar el modelo. Se han realizado análisis estadísticos adicionales utilizando la literatura existente para determinar los patrones de pronóstico de COVID-19 en relación con la prevalencia de metodologías, lenguajes de programación y fuentes de datos.

Conclusiones: este estudio de revisión puede ser valioso para investigadores, especialistas y tomadores de decisiones interesados en la administración de la pandemia del Virus Corona. Puede ayudar a desarrollar modelos de pronóstico y estrategias mejorados para la gestión de pandemias.

Palabras clave: Pronóstico; COVID-19; Predicción de Series Temporales; Aprendizaje Automático; Modelos Matemáticos; Modelos Basados en Datos.

INTRODUCTION

COVID-19 is a worldwide pandemic that has occurred in the 21st century. Initially documented in late 2019, this infectious disease quickly disseminated worldwide, with reported cases emerging from 228 countries. [1] Approximately 772 million individuals have been affected with this disease, resulting in approximately 6,9 million fatalities. [2] Due to its distinctive epidemiological characteristics, rapid transmission, and lack of viable therapy, social distancing, lockdowns, and quarantine measures have been implemented globally to restrict the spread of the disease and alleviate strain on healthcare systems. [3] The process of making vital determinations, such as whether to initiate the operation of colleges, schools, places of worship, and workplaces, relies heavily on accurate forecasts of the long-term prevalence of COVID-19. Providing explicit instructions and restrictions to citizens can be beneficial in reducing the spread of sickness. It is crucial to conduct a comprehensive analysis of the impact of these restrictions on other economic indicators, such as distress and joblessness. [4] Predicting the occurrence of COVID-19 can also assist in the formulation of treatment and vaccine plans, as well as in selecting high-priority areas for policymakers to concentrate their efforts on.

In order to effectively utilise the aforementioned strategies of regulating the spread, it is imperative to have a precise forecasting strategy. Over the past four years, a significant amount of material has been generated to support the necessity of anticipating COVID-19 trends. This literature takes into account various aspects of the disease and different geographical locations. Scientists have employed a diverse range of algorithms and models to successfully forecast the spread of COVID-19. Each of these approaches has produced varying levels of accuracy when applied to different datasets that encompass diverse geographical areas. In order to have a better understanding of the whole COVID-19 forecasting Paradigm, the present literature review focuses on the most often utilised algorithms and their respective study areas. This study provides a concise explanation of the literature search approach, followed by a quick overview of recent reviews on COVID-19. The subsequent sections have provided a concise overview of well-known forecasting models and algorithms. Ultimately, this paper presents a comprehensive analysis and evaluation of multiple published works, accompanied with statistical deductions and discussions. To the best of our knowledge, this review paper comprehensively examines a wide range of methodologies employed for COVID-19 forecasting, encompassing Mathematical, Statistical, Artificial Intelligence - Machine Learning, Ensembles, Transfer Learning, and hybrid approaches.

Scope of the Survey and Contributions

The purpose of this review is to provide a comprehensive explanation of the utilisation of disease forecasting technologies in predicting the COVID-19 pandemic. This study examines and analyses a comprehensive compilation of research papers published during the past four years. Its objective is to predict the trends of the COVID-19 pandemic, while providing guidance to policy makers on implementing steps to control the spread and avoid the disease.

To be specific, this review has been conducted in order to answer the following questions:
- Identify the most recent forecasting methodologies employed in the prediction and forecasting of the COVID-19 pandemic.
- Quantify and demonstrate the effectiveness and precision of the identified forecasting methods.
- Among scholars, which methodology is more renowned for forecasting the spread of infectious diseases?

- Which particular dataset is extensively used?
- Which performance measurements are commonly used?
- Which programming language has been used predominantly by researchers?
- Which geography is mostly covered by scholars while forecasting COVID-19 trends?

These questions are explored in this study, with the following major contributions:

- This paper represents the initial comprehensive analysis that specifically examines the wide-ranging topic of COVID-19 forecasts utilising all available methodologies. Several surveys focus on specific domains such as AI or Mathematical approaches for COVID-19 Forecasting. However, in this initial survey, a comprehensive methodology is employed to examine all literature pertaining to COVID-19 Forecasting, encompassing various methodologies and methods.
- Unlike previous similar studies, this one provides a comprehensive background description of every key technique used in the literature. This will unquestionably assist a non-expert in comprehending and assimilating the essential concepts.
- Every research study under analysis contains a comprehensive assessment of the rationale behind the chosen technique, emphasising the approach, dataset, geographic scope, performance metrics, and achieved results.
- Table 2-6 presents a comprehensive summary of the research literature that was examined on the forecasting techniques employed to predict the occurrence of COVID-19. This will enhance readers' ability to quickly scan through each literary work and grasp its significance in predicting COVID-19 outcomes.
- A comprehensive statistical analysis is conducted based on the chosen literature review, taking into account important criteria such as the popularity of Methods and Datasets being used, Language Preference, COVID-19 Features being investigated for Prediction, and Demographics Chosen for Study. This simplifies the process of comprehending the optimal techniques employed in forecasting approaches and the relevant factors that need to be considered for future endeavours in this domain.

Search methodology
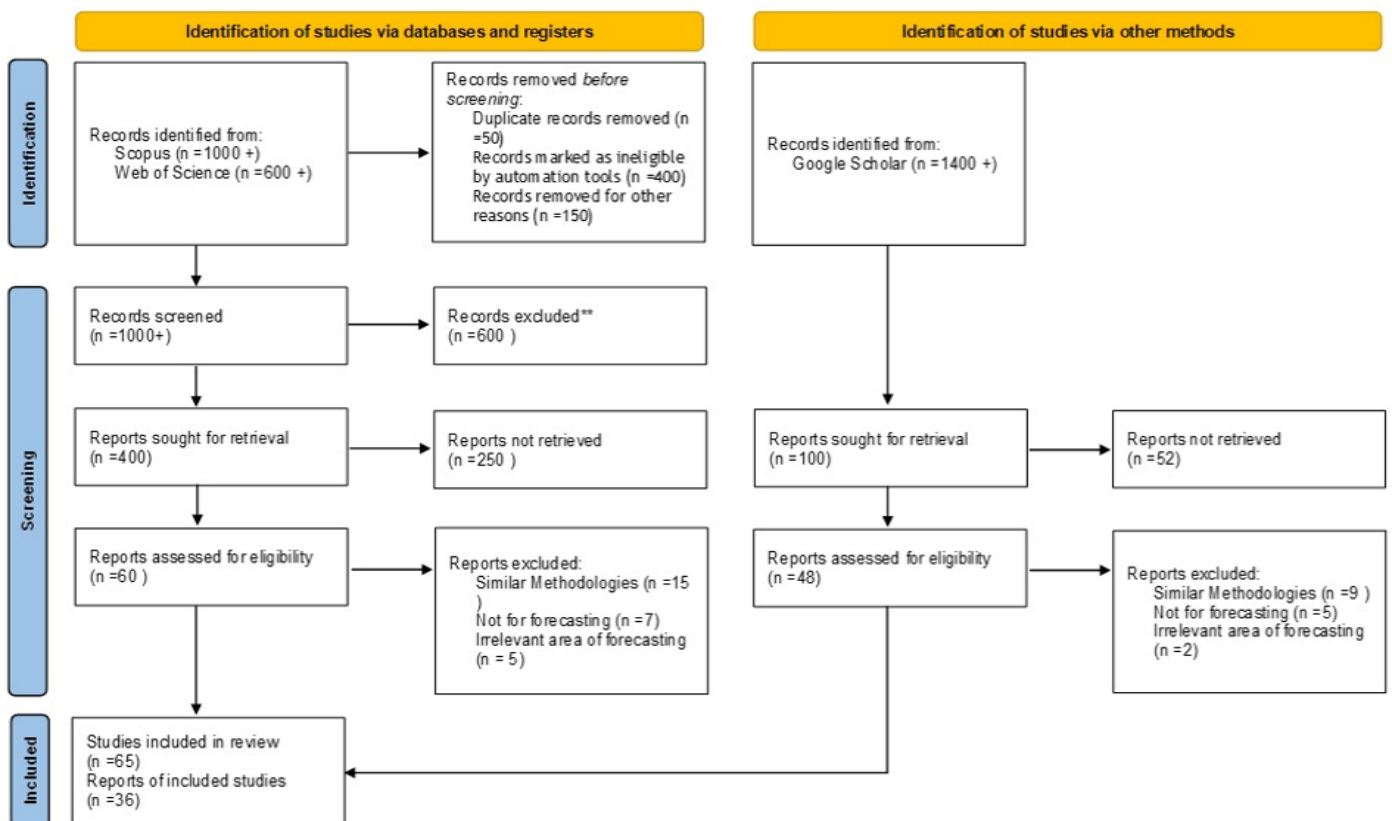
Figure 1. Pictorial Representation of Search Workflow Employed

The search methodology of linked material is crucial for the effectiveness of any review. An effective and enlightening review can only occur when the appropriate and pertinent literature is gathered and examined. This survey study specifically utilised the Web of Science, Scopus, and Google Scholar repositories to gather relevant material. A keyword-based search is used to get pertinent Articles, Research Papers, Conference

Papers, case studies, and review articles. The terms utilised are COVID-19, Corona Virus, Forecasting, and Prediction. The keywords were segregated using "OR" and "AND" operators to filter out pertinent content.

Figure 2. Network File analysis created by single keywork COVID-19

Figure 3. Network file with both COVID-19 and Forecast as the keywords

Keywords are indispensable for locating pertinent literature in a specific research field. Keywords, as commonly acknowledged, "represent the fundamental investigation of a study." A keywords network provides a visual depiction of an information area, showcasing the themes given and their interrelationships or organisation. The VOSviewer software generates a term co-occurrence network, while bibliographic information is acquired from Scopus. The search was conducted hierarchically, beginning with the most frequently used keyword, which in our case is COVID-19. This search has yielded almost 3000+ search results in terms of the number of literature matches. The network diagram obtained is displayed in figure 2.

In the second phase, the search was conducted by incorporating an additional keyword using the AND operator. This ensures that only literature containing both "COVID-19" and "Forecast" as keywords will be retrieved. This has decreased the quantity of associated publications to ~2000. The network file is depicted in figure 3. The network displayed in figure 3 for coronavirus forecasting consists of 652 nodes and 46 445 links. In the network visualisation, thicker lines indicate stronger connections. Figure 3 shows that the most commonly used search phrases by researchers were Coronavirus disease 2019, Coronavirus outbreak, computer science, timeseries, artificial intelligence, statistics, outbreak, and medicine. Figure 3 represents a cluster, which is shown by its colour. The keyword that is used more frequently is represented by a larger circle.

Given that this evaluation specifically examines COVID-19 Pandemic Forecasting, certain keywords such as Hospitality, Tourism, Finance, Economics, etc., are not particularly relevant. Next, the literature is further filtered by excluding irrelevant keywords from the search, resulting in a literature count ~1000. Figure 4 displays the network diagram of the same.

Figure 4. Network File after dropping not so relevant keywords related to COVID-19 Forecasting

Other related reviews

As of the time this study was written, a number of studies had reviewed the methodologies suggested for predicting the COVID-19 Pandemic's progression using a variety of techniques, including mathematical,

statistical, and AI-based models. The majority of the linked evaluations focus on one specific category of techniques that are utilised for COVID-19 predictions while covering several geographies. A summary of the most important and current reviews is given in the paragraphs that follow.

In a survey report on forecasting Coronavirus illness, Gitanjali et al.[5] divided forecasting methods into two main groups: data science/machine learning methods and mathematical models based on stochastic theory. They have assessed data from social media communications as well as big data collected from national and World Health Organization databases. They have assessed a number of criteria, including the effect of environmental conditions, the incubation period, the influence of quarantine, age, and gender, for the pandemic forecasting. In their review work on COVID-19 forecasting models, Rahimi et al.[6] primarily focused on machine learning

techniques. On the COVID-19 data, they also performed a bibliometric analysis, which was followed by a thorough categorization of machine learning models that included a comparison of approach and evaluation of the criteria.

Comito et al.[7] have carefully researched the usage of artificial intelligence-based techniques in their focused review on the COVID-19 Pandemic. Their research offers a thorough analysis of the approaches, formulas, tools, software, and latest AI methodologies employed to forecast and identify corona virus pandemic. Yi et al.[8] have assessed the COVID-19 system for artificial intelligence-based disease prevention and control. According to their review, the COVID-19 huge data, intelligent systems, and intelligent robots have all been used in the fight against the pandemic. This article thoroughly examined the uses and difficulties of AI technology during the pandemic, which is extremely important to further AI development and can act as a new standard for emergencies in the future.

Using information from the WHO and social media interactions, Mahalle et al.[9] classified forecasting approaches as mathematical models and machine learning techniques. Important factors like the number of deaths, metrological parameters, the length of the quarantine, the availability of medical services, and mobility were also examined.

In their review of COVID-19, Luo[10] suggested of using a heuristic approach and an experimental attitude, to tackle the challenge of unknown-unknown to handle the great uncertainty of the pandemic. To create governmental policies, organisation and planning, and individual mindset heuristically future-informed despite the severe uncertainty, he suggests and supports the "predictive monitoring" paradigm, which combines prediction with monitoring.

In order to assist COVID-19 research, Dagliati et al.[11] survey focuses on collaborative data infrastructures. The authors emphasised the proposed method in sharing and privacy of data, its rules and governance, and they drew attention to the issue interoperability of data caused by the variance in formats and standards of underlying data, modelling and exemplification of healthcare processes, and shared procedures. A review of the AI and clinical information systems based methodologies is the subject matter of study in Combi et al.[12]'s article. For COVID-19 data-intensive applications, the authors developed a nomenclature based on procedures and approaches for categorising judicious information routines and AI methods. The major characteristics of the functions, such as dataset collection, machine learning algorithm, use of natural language processing, data mining and trail identification, decision support systems, are outlined in the article in accordance with this taxonomy. Comparing their survey to earlier studies, Combi et al. survey 's was a little more technical, focusing mostly on computer science-related bibliographical sources.

To combat the COVID-19 pandemic, Anjum et al.[13] analysed a number of real-time and intelligent COVID-19 forecasting, diagnosing, and monitoring systems. They had created a taxonomy in their work that was based on the sophisticated COVID-19 predicting, diagnosing, and monitoring systems. They have examined a sizably broad spectrum of IoTs and machine learning techniques that can be used to diagnose and track infected people as well as anticipate the spread of COVID-19. The Potential of Artificial Intelligence Approaches to Forecasting COVID-19 Spreading has been reviewed by Jamshidi et al.[14] Their review attempts to summarise and assess the reliability and effectiveness of some significant AI-powered techniques employed for COVID-19 outbreak prediction. Review included 65 preprints, peer-reviewed papers, conference proceedings, and book chapters that were released in 2020. The findings showed that while the methods outlined in their review paper have a reasonable chance of accurately forecasting the spread of COVID-19, there are still some flaws and shortcomings that call for additional development and improvement.

A review of the deep learning techniques utilised for Covid-19 predictions was conducted by Kamalov et al.[15] For classifying the literature, they offered a model-based taxonomy and went on to further characterise each model and its performance. The SIR, SEIR, SEIRS, SIRD, and ARIMA, as well as cutting-edge data-driven approaches like the LSTM, AE, MAE, CNN, and Random Forest, have both been described in detail in Pereira et. al.[16] work's on pandemic forecasting. In comparison to conventional techniques, their review has outlined the primary benefits and limitations of data-driven strategies.

This research presents a comprehensive perspective on the methods used to predict COVID-19, which sets it apart from past studies on the subject. The present review encompasses widely recognised approaches for

forecasting pandemics: conventional mathematical models, data-driven techniques (utilising both statistical and machine learning), deep learning methods, Hybrid Methods, Ensemble and Transfer Learning based methods. Due to our focus on a diverse set of approaches, the evaluation offers a comprehensive and detailed overview of all methods, providing a deep understanding to those with expertise. However, the aforementioned evaluations concentrate on a particular domain and provide a general outline of the key methodologies, rather than delving into the extensive range of approaches discussed in the literature. The present study additionally incorporates a comprehensive bibliographic analysis of the COVID-19 literature accessible from Web of Science and Scopus. Furthermore, it analyses the statistical implications found in the literature being examined, offering readers a distinct perspective on the information.

In the next section, the methods utilised by researchers to study the COVID-19 pandemic dynamics including forecasting and the evaluation indices employed to gauge the outcomes produced are comprehensively described before a complete analysis of the publications reviewed in this survey is given.

METHOD

Researchers have considered a number of strategies and methods to predict the COVID-19 pandemic. The solution strategies / methods put out by researchers for forecasting COVID-19 are shown in table 1. Broadly these methods have been divided into Mathematical Models, Statistical Methods, AI and Machine Learning Methods and Hybrid Methods.

Table 1. Methods and Models for COVID-19 Pandemic Prediction			
Mathematical	**Statistical**	**AI and Machine Learning**	**Hybrid Methods**
Susceptible, Infected, and Recovered (SIR) Susceptible, Exposed, Infected, Recovered (SEIR) Susceptible, Infected, Recovered, Death (SIRD) Susceptible, Infected, Recovered, vaccinated (SIRV) Susceptible, Infected, Recovered, vaccinated, death (SIRVD) Singular Spectrum Analysis (SSA)	Auto Regression Moving Averages Weighted Moving Average (WMA) Simple Exponential Smoothing Holt Winters Exponential Smoothing Simple exponential smoothing with multiplicative error (ETS(MNN) Heterogeneous Autoregressive (HAR)	Adaptive neuro-fuzzy inference system (ANFIS) Generative Adversarial Network (GAN) Stacked Long short term Memory Networks (sLSTM) Bidirectional Long short term Memory Networks (BiLSTM) Multi linear regression Support Vector Regressor (SVR) Random Forest Artificial Neural Network (ANN) Multilayer perceptron (MLP) Polynomial Neural Network (PNN) Least Absolute Shrinkage and Selection Operator (LASSO)	Genetic Programming Flower Pollination Algorithm Ecological Niche models TCN-GRU-DBN-Q-SVM model Theta-ARNN (TARNN) Heuristic optimization algorithm (SCWOA) Ensemble Gaussian process regression (GPR) FTSOAX - improved seagull optimization algorithm (ISOA) and XGBoost Honey Badger Algorithm with Multilayer Perceptron adaptive neuro-fuzzy inference system-reptile search algorithm (ANFIS-RSA)

In subsequent sections, a brief description of these methods is provided.

Mathematical Models
Epidemic Compartmental Models

These are fundamental mathematical models for infectious diseases that spread through interaction between individuals within a community. The most basic models are expressed as initial value problems for a system of ordinary differential equations and are examined mathematically. All individuals of a given population, based on their state with respect to infection or disease, are divided into mutually exclusive compartments. It is assumed that one person will belong to one compartment based on its illness state. With time the person can move to other compartment if his illness state gets changed. This movement is determined by model parameters for compartment movement. The proposed models vary based on the number of compartments. The determination of the number of compartments is based on several criteria, including disease-related characteristics such as the infectious agent, method of transmission, latent period, infectious period, susceptibility and resistance, as well as social factors, demographics, cultural factors, and ecology.

SIR Model: this is the first and basic mathematical model for infectious disease modelling. This model is named on the basis of three different population divisions or compartments that it defines viz. susceptible, infected, and recovered. Every person in the population first starts from susceptible state from where it can move to infected state depending on his encounter with the infection and this determined by infection rate. After a period of being sick, the person transitions to a noncontagious condition called recovery. During the time they are infected, they are assumed to be capable of spreading the infection. However, this stage can also result in death or successful isolation.

SEIR Model: this model is enhanced version of SIR model where we have introduced another compartment of Exposed individuals. Exposed group of individuals are those members of the population which are being exposed to the virus but are still not contagious. This compartment falls between the Susceptible and Infected. The model uses below nomenclature to define its compartments. S - Susceptible, E - Exposed, I - Infected and R – Recovered.

The Recovered compartment contains both kind of individuals who are either recovered from the infection or succumbed to the infection. The model further defines a set of variables which controls the movement of individuals in between these states. These are defined below.

- B: is the Infection rate. The rate at which the infection spreads in a population. Hence this regulates the movement of individual from susceptible to infected compartments.
- δ: is the Incubation rate. The metric quantifies the percentage of susceptible individuals that acquire an infection.
- γ: is the Recovery rate. It quantifies the rate at which individuals infected with a disease regain their health.

The differential equations are given as below. Rate of change of susceptible group.

$$\frac{dS}{dt} = \beta \times S(t) \times I(t) \qquad (1)$$

Rate of change of exposed group.

$$\frac{dE}{dt} = \beta \times I(t) - \delta \times E(t) \qquad (2)$$
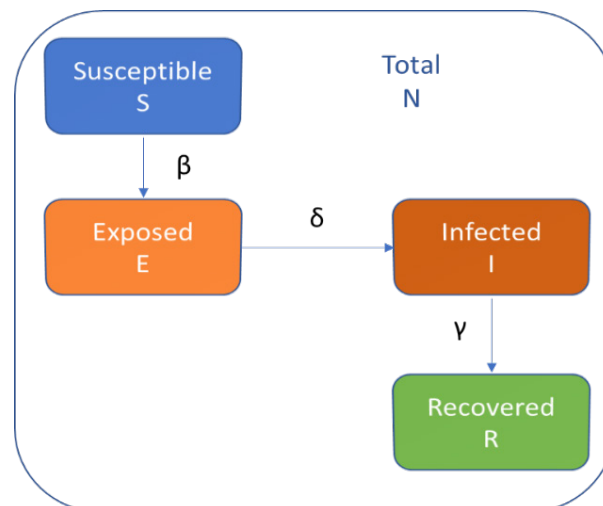
Figure 5. SEIR Model with compartments and parameters

Rate of change of infectious group.

$$\frac{dI}{dt} = \delta \times E(t) - \gamma \times I(t) \qquad (3)$$

Rate of change of recovered group.

$$\frac{dR}{dt} = \gamma \times I(t) \qquad (4)$$

SIRD Model: the SIRD model adds one more compartment in SIR model viz. Death. Basically, it further divide Recovered compartment of SIR model into individuals who recovered from the disease and individuals who succumbed from the infection. Hence it adds a new parameter μ which is defined as mortality rate. Due to this the differential equation gets changed to:

$$\frac{dS}{dt} = \beta \times S(t) \times I(t) \qquad (5)$$

$$\frac{dI}{dt} = \beta \times S(t) \times I(t) - \gamma \times I(t) - \mu \times I(t) \qquad (6)$$

$$\frac{dR}{dt} = \gamma \times I(t) \qquad (7)$$

$$\frac{dD}{dt} = \mu \times I(t) \qquad (8)$$

SIRV Model: one new compartment which accounts for vaccination of the susceptible population is added in SIR model to form SIRV model. It also introduces one more parameter – rate of vaccination denoted by ϑ. Due to this the differential equation gets changed to:

$$\frac{dS}{dt} = \beta \times S(t) \times I(t) - \vartheta(t) \times S(t) \qquad (9)$$

$$\frac{dI}{dt} = \beta \times S(t) \times I(t) - \gamma \times I(t) \qquad (10)$$

$$\frac{dR}{dt} = \gamma \times I(t) \qquad (11)$$

$$\frac{dV}{dt} = \vartheta \times S(t) \qquad (12)$$

SIRVD Model: the SIRVD model is an extension of the SIR model with an addition of two compartments one which accounts for vaccination of the susceptible population and the other death count. It also introduces two more parameters – rate of vaccination denoted by ϑ and rate of death due to disease as μ. Due to this the differential equation gets changed to:

$$\frac{dS}{dt} = \beta \times S(t) \times I(t) - \vartheta(t) \times S(t) \qquad (13)$$

$$\frac{dI}{dt} = \beta \times S(t) \times I(t) - \gamma \times I(t) - \mu \times I(t) \qquad (14)$$

$$\frac{dR}{dt} = \gamma \times I(t) \qquad (15)$$

$$\frac{dV}{dt} = \vartheta \times S(t) \qquad (16)$$

$$\frac{dD}{dt} = \mu \times I(t) \qquad (17)$$

Singular Spectrum Analysis (SSA)

A new and effective method for time series analysis, singular spectrum analysis (SSA) combines dynamical systems, signal processing, multivariate statistics, and multivariate geometry with traditional time series analysis. The goal of SSA is to break down the original series into a handful of distinct, comprehensible elements, such as a slowly changing trend, oscillatory elements, and random noise. SSA has a wide range of potential applications, including those in physics, economics, meteorology, financial mathematics, social science, and market research.

An essential characteristic of the SSA decomposition is that the original time series must meet a linear recurrent formula (LRFs). In order to forecast the new data points, the SSA approach should often be used with time series controlled by linear recurrent formulas. The empirical method and the bootstrap method are the two ways to construct confidence intervals based on the SSA technique. The complete series' empirical confidence intervals are created, and it is expected that the future structure will remain the same. For the continuation of the signal, which constitutes the major elements of the entire series, bootstrap confidence intervals are constructed.

So basically, SSA is a non-parametric method for estimating spectral properties by incorporating a time series into its analysis. {X(t):t=1,2, ..., N} in a vector space of dimension M. The Singular Spectrum Analysis (SSA) method involves the process of diagonalising the M x M lag-covariance matrix C_x of the time series X(t). This allows for the extraction of spectrum information from the time series, which is considered to be stationary in the weak sense. The matrix C_x can be easily predicted from the data as a Toeplitz matrix with constant diagonals. In other words, the entries c_{ij} of C_x solely depend on the lag $|i - j|$.

$$c_{ij} = \frac{1}{N-|i-j|} \sum_{t=1}^{N-|i-j|} X(t)X(t+|i-j|) \qquad (18)$$

Statistical Models
Heterogeneous Autoregressive Models (HAR)
Heterogeneous autoregressive models of realised volatility are widely used in financial market research, based on high-frequency measures of volatility and the idea that traders with different time horizons respond to and generate different volatility components. This basic model, estimated using conventional least squares regression, is compared favourably to other methods in equities, fixed income, and commodity markets.

The HARCH (heterogeneous autoregressive conditional heteroscedastic) model and the HAR-RV (heterogeneous autoregressive-realized volatility) model are two examples of studies of heterogeneous market hypotheses. Theoretically equal to an autoregressive model of order 22, using daily, weekly, and monthly moving averages, the HAR-RV model is a linear autoregressive model. The HAR-RV model incorporates both realised volatility and the heterogeneous market hypothesis. The heterogeneity it depicts spans the daily, weekly, and monthly timeframes. Long-memory is one of the HAR model's key characteristics. In precise, it excels in predicting realised volatility and accurately captures the properties of changeability, such as long-term memory.

There exist two categories of HAR models: univariate HAR models and bivariate HAR models. These are described as below.

A (univariate) HAR model $\{Xt, t \in \mathbb{Z}\}$ of order p is given by:

$$X_t = \emptyset_1 X_{t-1}^{(1)} + \emptyset_2 X_{t-1}^{(2)} + \cdots + \emptyset_p X_{t-1}^{(p)} + \in_t \qquad (19)$$

Where:

$$X_{t-1}^{(i)} = \frac{1}{h_i}\left(X_{t-1} + \cdots + X_{t-h_i}\right)$$

i = 1, 2, ..., p with positive integers $\{h_i, I = 1, 2, ..., p\}$ satisfying $1 = h_1 < h_2 < \cdots < h_p$ and $\{\in_t\}$ is a sequence of random variables with mean zero and variance $\sigma_\in^2$.
$\emptyset_1, ..., \emptyset_p$ are coefficients of the model to be estimated and it is assumed that:

$$\sum_{i=1}^{p} \emptyset_i < 1$$

A (bivariate) HAR model $\{(Xt, Y_t), t \in \mathbb{Z}\}$ of order (p, q) is defined as:

$$X_t = \alpha_{11} X_{t-1}^{(1)} + \cdots + \alpha_{1p} X_{t-1}^{(p)} + \beta_{11} Y_{t-1}^{(1)} + \cdots + \beta_{1q} Y_{t-1}^{(q)} + \in_{1,t} \qquad (20)$$

$$Y_t = \alpha_{21} X_{t-1}^{(1)} + \cdots + \alpha_{2p} X_{t-1}^{(p)} + \beta_{21} Y_{t-1}^{(1)} + \cdots + \beta_{2q} Y_{t-1}^{(q)} + \in_{2,t} \qquad (21)$$

Where: Y_{t-1}^i is given in the same way with h_i, i = 1, 2, ..., q; $\in_{1,t}$ and $\in_{2,t}$ are independent noise processes with mean zeros. Coefficients α_{ji}, β_{jk} are assumed to be:

$$\sum_{i=1}^{p} \alpha_{ji} + \sum_{k=1}^{q} \beta_{jk} < 1; j=1, 2$$

AI Models
Adaptive Neuro-Fuzzy Inference System (ANFIS)
The combination of artificial neural network (ANN) with principles of fuzzy logic gives birth to adaptive neuro-fuzzy inference system "ANFIS". Based on the Takagi-Sugano fuzzy inference system, it was created by him in the early 1990s. AFIS has the potential to capture the benefits of both neural networks and fuzzy logic principles in one framework. It is an adaptive network-based fuzzy inference system with learning capabilities to approximate nonlinear functions. AFIS is considered a universal estimator and used in intelligent situation-aware energy management systems. There are five design layers in ANFIS, which are: fuzzification layer, rule layer, normalization layer, result layer, and defuzzification layer. ANFIS is used for various applications such as pattern recognition, image recognition, facial recognition, and data mining.

Generative Adversarial Network (GAN) and GAN-GRU
GAN is a composite generative model. GAN uses two deep learning models called Generative (G) and Discriminative (D). The G model generates fake (or noisy) data to feed into D. The D model learns to distinguish

between fake and true data. By consecutively updating the constraints of generator "G" and discriminator "D" the GAN is trained. At the end of GAN training, the data distribution learned by G should be similar to the data distribution learned by D. The fundamental concept of GANs relies on the utilisation of "indirect" training through the Discriminator. Discriminator is a neural network capable of discerning the level of "realism" in an input and is continuously updated. Further the generator is not precisely taught to minimise the distance to a certain image, rather it is trained to deceive the discriminator. This permits the method to learn unsupervised. GAN-GRU and GAN-DNN are variants of GAN, where the first variant uses Generative and Discriminator as GRUs, and the second variant uses Deep Neural Network.
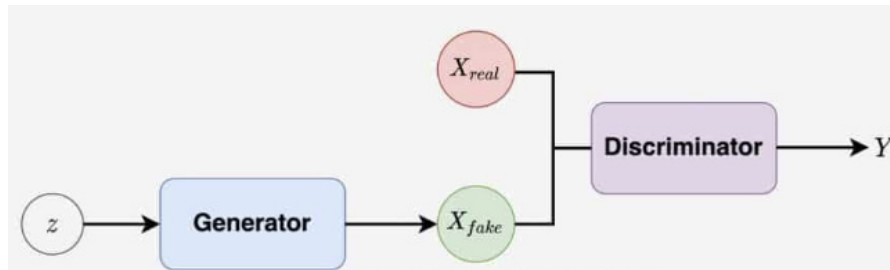
Figure 6. GAN Architecture

Polynomial Neural Network (PNN)

Polynomial Neural Networks (PNNs) are neural architectures that are adaptable and whose structure is generated through learning. Specifically, the PNN expands during the training period because its number of layers is dynamic rather than fixed in advance. PNN is a self-organizing network in this way. PNNs are very flexible because every node (processing element that forms a partial description) can use a different order of the polynomial (e.g., cubic, linear, quadratic, etc.) and have a different number of input variables. Unlike well-known neural networks, whose topologies are typically determined before any in-depth (parametric) learning occurs, the PNN architecture is fully optimised (both structurally and parametrically) without any prior fixation. In particular, the PNN architecture's layer count can be changed and additional layers added as needed.[17]

The PNN algorithm uses a class of polynomials that includes cubic, linear, modified quadratic, and other types. It is based on the GMDH approach. We may select the best forms from the recovered PDs based on both selecting nodes of each layer by selecting the most important input variables and an order for the polynomial among these different kinds of forms. Up until the expanded model reaches its peak performance, more layers are created. Such a process results in the best possible PNN structure. Recall that the input-output data are provided in the following format:

$$(X_i, y_i) = (x_{1i}, x_{2i}, \dots, x_{Ni}, y_i) \qquad (22)$$

Where: i = 1, 2, 3 … n
The input–output relationship of the above data by PNN algorithm can be described in the following manner:

$$y = f(x_1, x_2, \dots, x_N) \qquad (23)$$

The estimated output ŷ reads as:

$$\hat{y} = \hat{f}(x_1, x_2, \dots, x_N) = c_0 + \sum_{k_1} c_{k1} x_{k1} + \sum_{k_1 k_2} c_{k1k2} x_{k1} x_{k2} + \sum_{k_1 k_2 k_3} c_{k1k2k3} x_{k1} x_{k2} x_{k3} + \cdots \qquad (24)$$

Where: c_k's denotes the coefficients of the model.

Least Absolute Shrinkage and Selection Operator (LASSO)

Least Absolute Shrinkage and Selection Operator; (also Lasso or LASSO) is a regression analysis method used in statistics and machine learning that combines regularisation and variable selection to improve the statistical model's predictability and interpretability. Despite being designed primarily for linear regression, lasso regularisation can be easily extended to other statistical models, such as M-estimators, generalised estimating equations, proportional hazards models, and generalised linear models. The form of the constraint determines Lasso's ability to execute subset selection; this ability can be interpreted in a number of ways, including geometric, Bayesian, and convex analysis contexts. Consider the Multilinear regression equation

$$y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon \qquad (25)$$

y = the predicted value of the dependent variable.
β_0 = the y-intercept (value of y when all other parameters are set to 0).
$\beta_1 X_1$ = the regression coefficient (ß1) of the first independent variable (X1).
$\beta_n X_n$ = the regression coefficient of the last independent variable.
ε = model error.
The values for ß0, ß1, B2, … , ßn are chosen using the least square method, which minimizes the sum of squared residuals (RSS):

$$RSS = \sum (y_i - \hat{y}_i)^2 \qquad (26)$$

However, multicollinearity may become an issue if the predictor variables have a high degree of correlation. This may lead to large volatility and unreliability in the model's coefficient estimates. In other words, the model is likely to perform poorly when it is applied to a new collection of data that it has never seen before.
To solve this issue lasso regression is used which instead seeks to minimize the following:

$$RSS + \lambda \sum |\beta_j| \qquad (28)$$

Where: j ranges from 1 to p and $\lambda \geq 0$.
This second term in the equation is known as a shrinkage penalty.

DISCUSSION

In this section, the detailed finding of review in terms of paper / literature being analysed is presented. We have divided the literature on the basis of the methods they have employed for COVID-19 Forecasting. While selecting the literature, we have tried to cover an extended range of methods used for forecasting and hence we have ignored the articles which have used similar methods.

Literature Review of Articles Using Mathematical and Statistical Methods for COVID Forecasting
This section covers the literature being survey which have used either Mathematical or Statistical methods for COVID-19 forecasting. The summary of papers selected for this section is captured in table 2. Sarkar et al.[18] proposed a mathematical model accurately forecasts the patterns and changes of COVID-19 in 17 provinces of India as well as the entire country. The model tracks six compartments: susceptible, asymptomatic, recovered, infected, isolated infected, and confined susceptible. An examination of sensitivity indicates that decreasing the rates of interaction between persons who are not sick and those who are affected can result in a reduction of the basic reproduction number. The correctness of the model is contingent upon the implementation of enforced quarantine, isolation, and preventative measures.
Kibria et al.[19] used various models to predict daily COVID-19 cases in Bangladesh from April 2021 to July 2021. ARIMA model predicted an increase in daily cases within four weeks if strict regulations were not enforced. Qaness et al.[20] developed a novel short-term forecasting model using an upgraded version of the adaptive neuro-fuzzy inference system and an enhanced marine predator's algorithm.
Friji et al.[21] developed a comprehensive mechanistic model for the COVID-19 pandemic, targeting eight states. The model's parameters were established using an optimization problem involving infected cases, deceased cases, and reported cases. The Levenberg-Marquardt algorithm was used to resolve the problem. The model was evaluated using COVID-19 data from four nations, demonstrating precision. Nkwayep et al.[22] introduced an Ensemble of Kalman filter (EnKf) methodology to estimate unobservable state variables and unknown parameters in a COVID-19 model. The study concluded that the disease will persist without intervention, with transmission from infected individuals being the primary factor.
Kalantari et al.[23] used machine learning techniques to predict the global pandemic, with the report predicting its highest point by January 2021. The Singular Spectrum Analysis (SSA) model was found to be a feasible choice for predicting daily confirmed cases, fatalities, and recoveries. Hwang et al.[24] used heterogeneous autoregressive (HAR) time series models to analyze COVID-19 time series data in South Korea, with the bivariate HAR model analyzing scenarios with a substantial correlation between confirmed and confirmed cases. The HAR model offers a reliable predictive framework for COVID-19. Overall, these studies provide valuable insights into the pandemic's transmission and potential outcomes.

Table 2. Literature Review of Articles Using Mathematical and Statistical Methods for COVID-19 Forecasting						
Reference	**Algorithm Used**	**Metrics Used**	**Dataset Used**	**Features Used**	**Studied Regions**	**What is Predicted**
Sarkar et al.[18]	Mathematical Model - SARIIqSq - susceptible (S), asymptomatic (A), recovered (R), infected (I), isolated infected (Iq) and quarantined susceptible (Sq)	Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)	www.who.int, health.delhigovt.nic.in, dhs.kerala.gov.in, www.wbhealth.gov.in, www.covid19india.org	Daily new COVID-19 cases and Cumulative COVID-19 cases Turning and Ending dates of COVID-19	India and its 17 states	The values of the parameters are determined by adapting a time series solution of the SARII q S q model with the observed daily new cases of COVID-19 using the least square technique. How to control - Reproduction Number - R0.
	Autoregressive integrated moving average- ARIMA (8,1,7) also compared with ARMA, AR, Rolling forecast origin and MA	RMSE, MAPE, R2, MAE	worldometer	Daily number of confirmed cases	Bangladesh	ARIMA model yielded best outcomes, forecasting an increase in daily cases within a span of four weeks. dataset is from Apr 20 to Jul 28, 2021, prediction till Aug 30 2021
Kibria et al.[19]	Adaptiveneuro-fuzzy inference system (ANFIS). chaoticMPA (CMPA -	RMSRE, MAE, MAPE, and RMSE	WHO	Total confirmed cases	Russia and Brazil	Dataset 26 March to 26 Oct 2020 and predicted for next two weeks. Concluded that CMPA has shown better outcomes.
	Mathematical Model-S, P(Protected), E, I, Q (Quarantined), H(hospitalized), D0	MSE, MAE, NRMSE	Official Data - Not disclosed	The number of infected, deceased, and reported cases	Russia, Brazil, Italy, and USA	Death, infected and recovered count prediction
Qaness et al.[20]	SIRV-V (Free coronaviruses concentration in the environment) with Ensemble of Kalman filter (EnKf)	MAE and RMSE	Not Disclosed	Daily new cases, Cumulative Cases, Recovered Cases, Daily Deaths	Cameroon	Calculation of R(0) and forecast of Daily new cases, Cumulative Cases, Recovered Cases, Daily Deaths, for Cameroon
	Singular Spectrum Analysis (SSA)	Root Mean Squared Error (RMSE),	Johns Hopkins University	Daily confirmed cases, deaths, and recoveries	USA, India, Brazil, Russia, France, Spain, etc.	Forecasting is done for three features for next 40 days (starting from Oct 30 - 2020) for all the top 10 countries affected by COVID.
Friji et al.[21]	Heterogeneous autoregressive (HAR) time series models	RMSE and mean absolute error (MAE), R2, AIC, and BIC.	https://coronaboard.kr/ and http://www.ecdc.europa.eu/en/covid-19-pandemic.	Cumulative confirmed, recovered, death case. Recovery rate, Fatality rate and Infection rates	South Korea	forecasting is carried out with the COVID-19 data sets to obtain multi-step ahead predicted values and 95 % prediction intervals.
	Modified SEIRD dynamic model and ARIMA models	MAE, MSE, MLSE, Normalized MAE, and Normalized MSE.	The COVID Tracking Project	infected, recovered, and deceased populations	USA	Exhibited precise forecasts for the number of individuals who were infected, recovered, and perished.
Nkwayep et al.[22]	Time Series Analysis (TSA), hybrid ARIMA	NA	Not Disclosed	Number of deaths, confirmed cases, and recovered individuals	NA	The algorithm forecasts future cases by examining patterns and worldwide shifts. hybrid ARIMA algorithm is suggested to enhance accuracy.

Ala'raj et al.[25] developed a dynamic hybrid model that combines the SEIRD model and ascertainment rate to analyze COVID-19 attributes. The model uses real-time data and generates long- and short-term forecasts with confidence ranges. This model can help government entities, private industries, and policymakers mitigate health and economic hazards. Nramban et al.[26] developed a machine learning model for COVID-19 diagnosis, focusing on early detection. Time Series Analysis (TSA) was used to examine the number of deaths, confirmed cases, and recovered individuals during the pandemic. A novel hybrid ARIMA algorithm was suggested to enhance accuracy.

Literature Review of Articles Using Machine Learning Methods for COVID Forecasting

This section will cover the review of articles which have used Machine Learning Techniques for COVID-19 predictions.

Dairi et al.[27] compared different machine learning approaches for forecasting Covid-19 transmission, finding that hybrid deep learning models are effective in accurately predicting cases. Zain et al.[28] developed a hybrid CNN-LSTM model to predict COVID-19 cases using a dataset with changes over time. The model achieved the lowest average MAPE, RMSE, and RRMSE values compared to 17 baseline models. Shukla et al.[29] proposed mathematical modelling and artificial intelligence systems to monitor disease progress and forecast future patterns. The SEIRS epidemic model uses an optimized LSTM model to forecast confirmed cases, recovered patients, and fatalities. Cross-validation ensures model performance. Morais et al.[30] used a hybrid forecasting model to enhance resource allocation in the context of Covid-19, capturing weekly seasonality using a neural network model.

Alali et al.[31] developed a proficient model for predicting COVID-19 instances in India and Brazil using Gaussian process regression, Bayesian optimization, and the Random Forest algorithm. Masum et al.[32] used a mathematical epidemic model (MEM), statistical model, and recurrent neural network (RNN) variations to predict COVID-19 cases. The results showed that RNN versions were more accurate than MEM in forecasting, but MEM still provided valuable insights into virus propagation and control. Mann et al.[33] used an enhanced neural network called XBNet to analyze coronavirus instances in India over 30 days. The XBNet model had a precision accuracy of 99,27 % and outperformed other models in terms of MSE, MAPE, RMSE, and MAE. Prakash et al.[34] used a data-driven methodology, analyzing weather conditions and air quality index to forecast COVID-19 propagation in India.

Literature Review of Articles Using Hybrid Methods for COVID Forecasting

In this section, we will cover the Articles which have used hybrid methods for COVID forecasting. Cinaglia et al.[35] used neural networks to predict R0 in a mathematical model, resulting in a better MAPE. Manohar et al.[36] used ANN models to predict the COVID-19 outbreak in India between January and October 2021, with the ANN-BP model showing superior performance. Kumar et al.[37] introduced a hybrid fuzzy time series model using modified fuzzy C-means clustering to forecast COVID-19 cases and fatalities in India. The model consists of two phases: basic interval formation and sub-interval upgrade, forecasting cases and deaths for the next 31 days and estimating isolation beds and ICUs.

The study by Jithendra et al.[38] aims to improve time-series modelling for COVID-19 cases by combining a machine-learning model with a nature-inspired Reptile Search Algorithm. The ANFIS-RSA strategy, which uses data from China and India, achieved high precision with an R2 value of 0,9775, demonstrating its effectiveness in predicting COVID-19 cases. Xian et al.[39] introduces a novel approach, called FTSOAX, for predicting fuzzy time series. It utilises an enhanced seagull optimisation algorithm (ISOA) and the XGBoost algorithm. ISOA divides the discourse domain into more appropriate intervals, while SOA enhances convergence using the Powell algorithm and random curve action. Additionally, XGBoost predicts changes in fuzzy membership. FTSOAX has superior performance compared to other fuzzy forecasting algorithms in accurately estimating the number of daily confirmed COVID-19 cases in seven countries.

Prakash et al.[40] employs sophisticated techniques such as LSTM, KNN, SVR, Random Forest, and Multi Linear Regressor, in addition to the mathematical model SEIR, to forecast the daily occurrence of COVID-19 in India. Based on the predictions of the majority of models, it is expected that there will be no fourth wave of the COVID-19 pandemic in India in the near future, after a period of 200 days. The firefly technique is suggested by Melin et al.[41] for optimising ensemble neural networks in the prediction of COVID-19 time series. This is achieved through the utilisation of type-2 fuzzy logic and weighted average integration. It utilises the collective responses of individual artificial neural networks to make a final prediction, which provides benefits compared to traditional average approaches and type-1 fuzzy weighted average integration.

			Table 3.	Literature Review of Articles Using Machine Learning Methods for COVID-19 Forecasting		
Reference	**Algorithm Used**	**Metrics Used**	**Dataset Used**	**Features Used**	**Studied Regions**	**What is Predicted**
Dairi et al.[27]	LSTM-CNN, GAN-GRU, GAN, CNN, LSTM, and Restricted Boltzmann Machine (RBM), (LR) and (SVR).	R2, RMSE, MAE, MAPE, EV, MSLE	John Hopkins University	Confirmed and recovered COVID-19 cases	Brazil, France, India, Mexico, Russia, Saudi Arabia, and the US	Superiority of the LSTM-CNN model demonstrated by its enhanced performance, achieving an average mean absolute percentage error of 3,718 %.
	CNN-LSTM	MAPE, RMSE, and RRMSE	WHO COVID-19 dashboard	Daily confirmed cases	Global	The CNN LSTM is compared with 17 different algorithms and finally taken as best of the basis of metrics values.
Zain et al.[28]	Optimized LSTM	R2, RMSE	WHO	Daily new cases of COVID	India	Forecasted confirmed cases from Jan to April 2022 in India
	ARIMA and Neural Network models	MASE SMAPE R2 Coefficient	Our World In Data (OWID)	Daily Covid-19 cases	world	Proposed a hybrid model using ARIMA and Neural Network for Linear and Nonlinear data series.
Shukla et al.[29]	Gaussian process regression (GPR)	(RMSE), (MAE), (MAPE)	John Hopkins	Daily confirmed and recovered COVID-19	India and Brazil	The proposed model demonstrated substantial enhancement and outperformed other models, attaining an average error rate of approximately 0,1 %
	MEM (Mathematical model), bidirectional RNN, RNN, GRU, LSTM, BRNN, ARIMA	RMSE, MAE, MAPE	Not Disclosed	Confirmed cases daily	USA	Deep Learning based methods are more accurate in predictions while Mathematical models help to infer potential control strategies of pandemic
Morais et al.[30]	Extremely boosted neural network (XBNet)	MAPE, RMSE, MAE, and MSE	COVID-19 in India (2019) Kaggle - https://www.kaggle.com/datasets/sudalairajkumar/covid19-in-india	Total daily corroborated infested cases, mortality rate, discharge rate, growth factor	India	Projected the prediction of the cumulative number of confirmed COVID-19 cases for various time periods of India
	SVR, KNN Regressor, Multilinear regressor and Random Forest Regressor	MAPE, RMSE, MAE, and MSE	COVID19 INDIA (https://www.covid19india.org/).34	Total Number of Positive cases, Active cases on daily basis, Daily Deaths, New cases	India	This study has developed a model that can accurately forecast the spread of COVID-19 cases and fatalities over a period of 180 days. Furthermore, it has the capability to establish a relationship between weather conditions and the Air Quality Index in order to determine their influence on the spread of COVID-19.

Table 4. Literature Review of Articles Using Hybrid Methods for COVID-19 Forecasting

Reference	Algorithm Used	Metrics Used	Dataset Used	Features Used	Studied Regions	What is Predicted
Cinaglia et al.[35]	NN with R(0)	MAPE, MAE, RMSE	Our World in Data	Daily new positive cases	Italy, USA, UK, Sweden	Better MAPE obtained with the proposed new model
	A Boltzmann Function-based and Beesham's prediction - ANN-BP model	MSE and MAPE	https://covid19.telangana.gov.in/	daily COVID-19 positive cases	India - AP - 6 districts	Daily new positive cases while considering the data from 01-Jan-2021 till 31-Oct-2021
Manohar et al.[36]	Fuzzy C-means clustering technique	RMSE, MSE and Average forecast error rate	https://www.covid19india.org	Daily infected cases and death cases	India	Prediction of Daily infected cases and death count for next 31 days
	adaptive neuro-fuzzy inference system-reptile search algorithm (ANFIS-RSA)	RMSE, RMSRE, MAE, and MAPE, R2	WHO	Daily infected number of COVID cases	India and China	To enhance the ANFIS model to make it more accurate in optimizing non-linear problems and time-series predictions.
Kumar et al.[37]	FTSOAX - improved seagull optimization algorithm(ISOA) and XGBoost	RMSE, SMAPE	John Hopkins University	daily cases	USA, India, UK, Russia, Iran, Norway, Japan	Accurately estimated the number of daily confirmed COVID-19 cases in seven countries.
	SEIR, LSTM, KNN, SVR, Random Forest, and Multi Linear Regressor	MAE RMSE	John Hopkins	Daily new cases	India	Possibility of Covid fourth wave in India, next 200 days prediction of daily new cases
Jithendra et al.[38]	Firefly algorithm, ensemble neural network optimization applied with type-2 fuzzy logic	MSE	Humanitarian Data Exchange (HDX)	Confirmed and death cases due to COVID	26 countries are analyzed	Forecasted daily COVID cases and death count for the selected countries. - Austria, Belgium, Bolivia, Brazil, China, Ecuador, Finland, France, Germany, Greece, India, Iran, Italy, Mexico, etc.

Literature Review of Articles Using Ensemble Methods for COVID Forecasting

In this section, we will cover the Articles which have used ensemble methods for COVID forecasting. Bhattacharyya et al.[42] proposed the Theta-ARNN (TARNN) model is a hybrid technique created by combining the Theta method and autoregressive neural network to forecast COVID-19. This model, which has a satisfactory prediction error rate, can assist healthcare and government in efficient planning and allocation of resources, surpassing conventional univariate and hybrid forecasting models on average in test datasets. Jin et al.[43] proposed TCN-GRU-DBN-Q-SVM model is a hybrid ensemble model that integrates Temporal Convolutional Networks (TCN), Gated Recurrent Unit (GRU), Deep Belief Networks (DBN), Q-learning, and Support Vector Machine (SVM) models. Its purpose is to accurately predict the number of COVID-19 infections. The model utilises weights obtained through reinforcement learning and incorporates an error predictor constructed using Support Vector Machines (SVM). The experimental results demonstrate that the model has a satisfactory ability to forecast, efficiently predict errors, and update weights based on Q-learning. This ensures accuracy, resilience, and generalisation across many countries and conditions.

Qu et al.[44] developed a hybrid sine cosine algorithm-whale optimisation algorithm for COVID-19 outbreak prediction, demonstrating the efficacy of neural network models. Ribeiro et al.[45] evaluated various models for time series prediction in Brazilian states, revealing superior performance in various settings. Shastri et al.[46] used three advanced deep learning models: Bi-directional LSTM, Convolutional LSTM, and the CoBiD-Net ensemble, achieving superior accuracy and mean absolute percentage error. Cramer et al.[47] developed a multimodal ensemble forecast for COVID-19 fatalities, but found that the precision decreased over longer time periods. Ke et al.[48] presented a novel approach for predicting daily new 2019 COVID-19 case time series, combining ensemble empirical mode decomposition, fuzzy entropy reconstruction, and a CNN-LSTM-ATT hybrid network model. This approach provides technological assistance for predicting future outbreaks of infectious diseases.

Table 5. Literature Review of Articles Using Ensemble Methods for COVID-19 Forecasting

Reference	Algorithm Used	Metrics Used	Dataset Used	Features Used	Studied Regions	What is Predicted
Bhattacharyya et al.[42]	Theta-ARNN (TARNN)	RMSE, MAE, MASE	Our World in Data	Daily new cases	USA, Brazil, India, the UK, and Canada	Forecast of COVID-19 cases for March 20-29, 2021, using proposed TARNN model
	TCN-GRU-DBN-Q-SVM	RMSE, MAE	Epidemic Intelligence team of ECDC.	Daily new cases	USA, India and UK	Comparison of different models on different dataset
Jin et al.[43]	A new type of heuristic optimization algorithm (SCWOA) Ensemble	MAPE, RMSE	John Hopkins	Daily new Cases, Total Cases, Daily Death count, Total Death Count	USA, India and Brazil	Comparison among models
	SVR and stacking ensemble	MAE, and symmetric MAPE	Brazilian State Health Offices	Cumulative confirmed cases	Brazil	Predicted cumulative confirmed cases for Brazilian states and established SVR, Stacking ensemble model as best.
Qu et al.[44]	Proposed ensemble of Bi- LSTM, Conv LSTM -CoBiD-Net.	MAPE	WHO	Number of confirmed cases and death cases	USA, India, Brazil	Predicted COVID parameters and shown excellent accuracy of Ensemble model.
	Ensemble	Weighted interval score, Relative MAE	https://covid19forecasthub.org/ and Zoltar forecast archive	Incident deaths	USA	Ensemble based model have shown better metrics for forecast as compared to baseline models.
Ribeiro et al.[45]	EEMD-FE-CNN-LSTM-ATT	R2, MAE, MAPE RMSE	World Health Organization	Daily Incidence data	USA, France, Russia	A new model having greater accuracy is presented for Infectious disease forecasting.

Literature Review of Articles Using Transfer Learning Methods for COVID Forecasting

In this section, we will cover the Articles which have used Transfer Learning methods for COVID forecasting. Roster et al.[49] work investigates the application of transfer learning in forecasting novel diseases in contexts with little data. It specifically focuses on utilising data from Brazil and evaluates the efficacy of several machine learning models in transferring information between dengue, Zika, influenza, and COVID-19 cases. The study employs both empirical and synthetic methodologies to assess the effectiveness of transfer learning. Chakraborty et al.[50] in their paper presents a data-driven transfer learning model for predicting COVID-19, taking into account the circumstances in four countries: the United States, Spain, Brazil, and Bangladesh. Four LSTM-RNN models were trained first and then further optimised using data specific to India. The model surpasses all other models in performance, properly forecasting daily cases by effectively managing various variables.

Table 6. Literature Review of Articles Using Transfer Methods for COVID-19 Forecasting						
Reference	Algorithm Used	Metrics Used	Dataset Used	Features Used	Studied Regions	What is Predicted
Roster et al.[49]	Transfer Learning, Random Forest, NN, RF with TrAdaBoost	None	Brazil dengue and influenzas data set for predicting Zika and COVID-19	Daily Covid cases	Brazil	The study employs both empirical and synthetic methodologies to assess the effectiveness of transfer learning.
	Ensemble RNN + LSTM and transfer learning	RMSE	NA	Daily Covid Casea	India, USA Spain, Brazil and Bangladesh.	Use of transfer learning for COVID predictions for India using four countries dataset and trained model as base.
Chakraborty et al.[50]	graph neural networks + transfer learning	Error	https://dataforgood.fb.com/tools/disease-prevention-maps/	Daily new cases of COVID	4 European countries	Used model to predict the number of COVID cases for each one of the regions for next 14 days
	attention-based RNN with Transfer Learning	MAPE	https://www.kff.org/coronavirus-covid-19/fact-sheet/coronavirus-tracker/.	confirmed cases per million people	Worldwide	Using Transfer learning predicted the COVID cases for 72 countries.
Panagopoulos et al.[51]	LSTM - Transfer Learning	MAE and RMSE	Our World in Data	Case count and death count	India, Germany, France, Brazil, Nepal	Model Trained on Italy and USA Covid datasets and used to predict for India, Germany, France, Brazil and Nepal.

Panagopoulos et al.[51] study employs graph representation learning techniques to forecast the number of COVID-19 cases by utilising graph neural networks. The model utilises nodes to represent areas, history, and edge weights in order to depict human migration. By employing a model-agnostic meta-learning approach, the method is evaluated in comparison to conventional forecasting methodologies in four European nations. The experimental results demonstrate the method's superiority, especially in secondary waves, with transfer learning yielding the most optimal model. Li et al.[52] proposed ALeRT-COVID model, which was trained on specific source nations, was applied to target countries. The model included a lockdown measure as a predictor and an attention mechanism to understand the impact of past cases. This transfer learning approach enhanced the accuracy of predictions in developing nations. Gautam et al.[53] used Transfer Learning in LSTM networks to forecast new COVID cases and fatalities, utilising models derived from early affected nations such as Italy and the US. The method's validity is assessed using data from Germany, France, Brazil, India, and Nepal.

Insights of Reviewed Literature on Other Parameters

The literature is further segregated on the basis of parameters like Popular Method Used, Programming Language Used, Popular COVID-19 datasets and Demographics Distribution. In this section we will capture the same.

As shown in figure 7a below, it is clear that AI & ML based methods are most popular and heavily used for predicting COVID-19. Almost half of the reviewed papers have used AI and ML based methods, followed by

Mathematical model at 22 %, statistical methods which contributed to 16 % of the papers. Hybrid methods are sparsely used in literature and contributes to 11 % only. In figure 7b below, similar statistics has been analysed for popular datasets that has been used in these literatures. It is evident from the analysis that researchers have used a wide array of dataset for their studies which is clear as ~50 % of the datasets are country specific or falling in others category. Rest 50 % have only used the well-known datasets like John Hopkins University (25 %), WHO (20 %), OWID (9 %), Worldometer (3 %).

A similar statistic has been captured for the popular programming language used by researchers and Python was the clear winner with 71 % followed by R at 21 % and MATLAB at 8 % as captured in figure 7c below. As captured in figure 7d below, it seems that most of the researchers 57 % have restricted their COVID-19 studies to country specific only, and this supports the COVID-19 spread and control measures in different geographies and its impact on datasets and trends. 43 % of studies were global in nature.
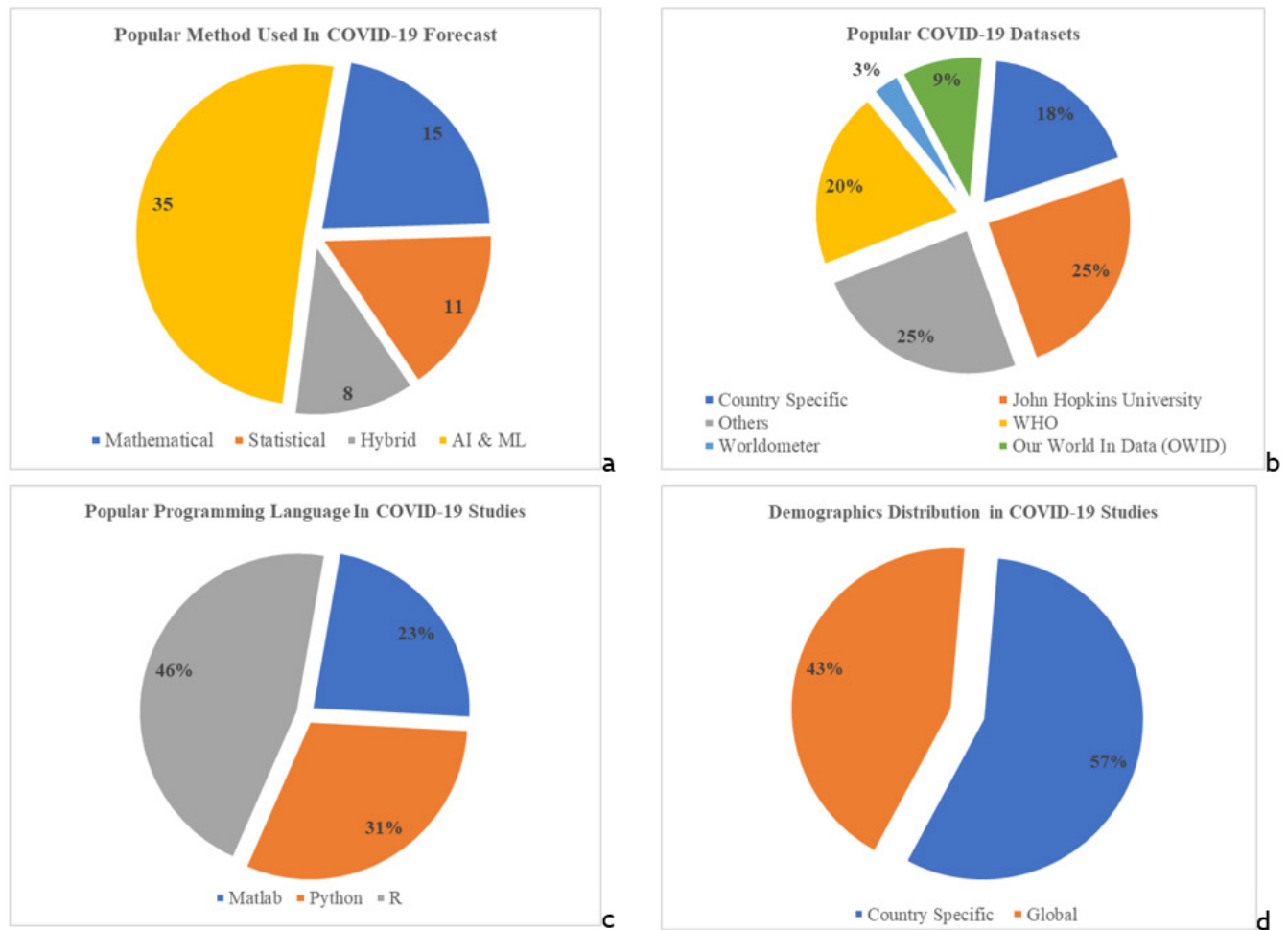
Figure 7. a) Popular Methods used in COVID-19 Forecast b) Popular COVID-19 datasets used c) Popular COVID-19 datasets used d) Demographics distribution of COVID-19 Studies

CONCLUSIONS

The report provided a methodical and thorough examination of the various techniques employed in the prediction of COVID-19. First of its kind, this review paper comprehensively examines a wide range of methodologies employed for COVID-19 forecasting, encompassing Mathematical, Statistical, Artificial Intelligence - Machine Learning, Ensembles, Transfer Learning, and hybrid approaches. The study also offers a thorough overview of the approaches employed for forecasting COVID-19 giving reader a good understanding of the employed approaches. A comprehensive analysis is conducted for each surveyed work, which includes a complete examination of the reasoning behind the approach. This study focuses on the methodology employed, the type and scale of data analysed, the method of validation, the intended application, and the realised outcomes. Although there has been notable advancement in utilising AI & ML and Mathematical based methods for COVID-19 forecasting, there is still a requirement for additional improvements in terms of precision and bolstering the implementation to effectively capture the COVID-19 waves and future trajectory.

BIBLIOGRAPHIC REFERENCES

1. COVID - Coronavirus Statistics - Worldometer [Internet]. [cited 2024 Sep 11]. Available from: https://www.worldometers.info/coronavirus/#countries

2. COVID-19 cases | WHO COVID-19 dashboard [Internet]. [cited 2024 Sep 11]. Available from: https://data.who.int/dashboards/covid19/cases?n=c

3. Chatterjee SC, Chatterjee D. COVID-19, Older Adults and the Ageing Society. Covid-19, Older Adults and the Ageing Society [Internet]. 2022 Jan 1 [cited 2024 Sep 11];1–155. Available from: https://www.taylorfrancis.com/books/oa-mono/10.4324/9781003286936/covid-19-older-adults-ageing-society-suhita-chopra-chatterjee-debolina-chatterjee

4. Rudolph CW, Allan B, Clark M, Hertel G, Hirschi A, Kunze F, et al. Pandemics: Implications for research and practice in industrial and organizational psychology. Ind Organ Psychol [Internet]. 2021 Jun 1 [cited 2024 Sep 11];14(1–2):1–35. Available from: https://www.cambridge.org/core/journals/industrial-and-organizational-psychology/article/abs/pandemics-implications-for-research-and-practice-in-industrial-and-organizational-psychology/1B702A23756307A6658F02576C8CED51

5. Shinde GR, Kalamkar AB, Mahalle PN, Dey N, Chaki J, Hassanien AE. Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art. SN Comput Sci. 2020 Jul 1;1(4).

6. Rahimi I, Chen F, Gandomi AH. A review on COVID-19 forecasting models. Neural Comput Appl. 2023 Nov 1;35(33):23671–81.

7. Comito C, Pizzuti C. Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review. Vol. 128, Artificial Intelligence in Medicine. Elsevier B.V.; 2022.

8. Yi J, Zhang H, Mao J, Chen Y, Zhong H, Wang Y. Review on the COVID-19 pandemic prevention and control system based on AI. Eng Appl Artif Intell. 2022 Sep 1;114:105184.

9. Mahalle P, Kalamkar AB, Dey N, Chaki J, ella Hassanien A, Shinde GR, et al. Forecasting Models for Coronavirus (COVID-19): A Survey of the State-of-the-Art. Authorea Preprints [Internet]. 2023 Oct 30 [cited 2024 Sep 11]; Available from: https://www.authorea.com/doi/full/10.36227/techrxiv.12101547.v1?commit=7514de8c8a55fb6aeb3b95b32dabf51a0d7cd24d

10. Luo J. Forecasting COVID-19 pandemic: Unknown unknowns and predictive monitoring. Vol. 166, Technological Forecasting and Social Change. Elsevier Inc.; 2021.

11. Dagliati A, Malovini A, Tibollo V, Bellazzi R. Health informatics and EHR to support clinical research in the COVID-19 pandemic: an overview. Brief Bioinform [Internet]. 2021 Mar 1 [cited 2024 Sep 11];22(2):812–22. Available from: https://pubmed.ncbi.nlm.nih.gov/33454728/

12. Combi C, Pozzi G. Health Informatics: Clinical Information Systems and Artificial Intelligence to Support Medicine in the CoViD-19 Pandemic. Proceedings - 2021 IEEE 9th International Conference on Healthcare Informatics, ISCHI 2021. 2021 Aug 1;480–8.

13. Anjum N, Asif A, Kiran M, Jabeen F, Yang Z, Huang C, et al. Intelligent COVID-19 Forecasting, Diagnoses and Monitoring Systems: A Survey [Internet]. Vol. 14, IEEE COMMUNICATIONS SURVEYS & TUTORIALS. 2021. Available from: https://ieeexplore.ieee.org/

14. Jamshidi MB, Roshani S, Talla J, Lalbakhsh A, Peroutka Z, Roshani S, et al. A Review of the Potential of Artificial Intelligence Approaches to Forecasting COVID-19 Spreading. Vol. 3, AI (Switzerland). Multidisciplinary Digital Publishing Institute (MDPI); 2022. p. 493–511.

15. Kamalov F, Rajab K, Cherukuri AK, Elnagar A, Safaraliev M. Deep learning for Covid-19 forecasting: State-of-the-art review. Neurocomputing. 2022 Oct 28;511:142–54.

16. Pereira IG, Guerin JM, Júnior AGS, Garcia GS, Piscitelli P, Miani A, et al. Forecasting covid-19 dynamics in brazil: A data driven approach. Int J Environ Res Public Health. 2020 Jul 2;17(14):1–26.

17. Oh SK, Pedrycz W, Park BJ. Polynomial neural networks architecture: analysis and design. Computers & Electrical Engineering. 2003 Aug 1;29(6):703–25.

18. Sarkar K, Khajanchi S, Nieto JJ. Modeling and forecasting the COVID-19 pandemic in India. Chaos Solitons Fractals. 2020 Oct 1;139.

19. Kibria HB, Jyoti O, Matin A. Forecasting the spread of the third wave of COVID-19 pandemic using time series analysis in Bangladesh. Inform Med Unlocked. 2022 Jan 1;28.

20. Al-qaness MAA, Saba AI, Elsheikh AH, Elaziz MA, Ibrahim RA, Lu S, et al. Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil. Process Safety and Environmental Protection. 2021 May 1;149:399–409.

21. Friji H, Hamadi R, Ghazzai H, Besbes H, Massoud Y. A generalized mechanistic model for assessing and forecasting the spread of the COVID-19 pandemic. IEEE Access. 2021;9:13266–85.

22. Nkwayep CH, Bowong S, Tewa JJ, Kurths J. Short-term forecasts of the COVID-19 pandemic: a study case of Cameroon. Chaos Solitons Fractals. 2020 Nov 1;140.

23. Kalantari M. Forecasting COVID-19 pandemic using optimal singular spectrum analysis. Chaos Solitons Fractals. 2021 Jan 1;142.

24. Hwang E, Yu SM. Modeling and forecasting the COVID-19 pandemic with heterogeneous autoregression approaches: South Korea. Results Phys. 2021 Oct 1;29.

25. Ala'raj M, Majdalawieh M, Nizamuddin N. Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections. Infect Dis Model. 2021 Jan 1;6:98–111.

26. Nramban Kannan SK, Kolla BP, Sengan S, Muthusamy R, Manikandan R, Patel KK, et al. Analysis of COVID-19 Datasets Using Statistical Modelling and Machine Learning Techniques to Predict the Disease. SN Comput Sci [Internet]. 2024 Jan 1 [cited 2024 Sep 12];5(1):1–18. Available from: https://link.springer.com/article/10.1007/s42979-023-02464-y

27. Dairi A, Harrou F, Zeroual A, Hittawe MM, Sun Y. Comparative study of machine learning methods for COVID-19 transmission forecasting. Vol. 118, Journal of Biomedical Informatics. Academic Press Inc.; 2021.

28. Zain ZM, Alturki NM. COVID-19 Pandemic Forecasting Using CNN-LSTM: A Hybrid Approach. Journal of Control Science and Engineering. 2021;2021.

29. Shukla SSP, Jain VK, Yadav AK, Pandey SK. Fourth wave Covid19 analyzing using mathematical seirs epidemic model & deep neural network. Multimed Tools Appl [Internet]. 2024 Mar 1 [cited 2024 Sep 12];83(9):27507–26. Available from: https://link.springer.com/article/10.1007/s11042-023-16609-x

30. de Araújo Morais LR, da Silva Gomes GS. Forecasting daily Covid-19 cases in the world with a hybrid ARIMA and neural network model. Appl Soft Comput. 2022 Sep 1;126:109315.

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

32. Masum M, Masud MA, Adnan MI, Shahriar H, Kim S. Comparative study of a mathematical epidemic model, statistical modeling, and deep learning for COVID-19 forecasting and management. Socioecon Plann Sci. 2022 Mar 1;80.

33. Mann S, Yadav D, Muthusamy S, Rathee D, Mishra OP. A Novel Method for Prediction and Analysis of COVID 19 Transmission Using Machine Learning Based Time Series Models. Wirel Pers Commun [Internet]. 2024 Dec 1 [cited 2024 Sep 12];133(3):1935–61. Available from: https://dl.acm.org/doi/10.1007/s11277-023-10836-z

34. Prakash S. A Robust Machine Learning Model for Prediction of COVID-19 Pandemic with Climate & Air Quality Parameters. [cited 2024 Sep 12]; Available from: https://aqicn.org

35. Cinaglia P, Cannataro M. Forecasting COVID-19 Epidemic Trends by Combining a Neural Network with Rt Estimation. Entropy. 2022 Jul 1;24(7).

36. Manohar B, Das R. Artificial neural networks for prediction of COVID-19 in India by using backpropagation. Expert Syst [Internet]. 2023 Jun 1 [cited 2024 Sep 12];40(5):e13105. Available from: https://onlinelibrary. wiley.com/doi/full/10.1111/exsy.13105

37. Kumar N, Kumar H. A novel hybrid fuzzy time series model for prediction of COVID-19 infected cases and deaths in India. ISA Trans. 2022 May 1;124:69–81.

38. Jithendra T, Sharief Basha S. A Hybridized Machine Learning Approach for Predicting COVID-19 Using Adaptive Neuro-Fuzzy Inference System and Reptile Search Algorithm. Diagnostics (Basel) [Internet]. 2023 May 1 [cited 2024 Sep 12];13(9). Available from: https://pubmed.ncbi.nlm.nih.gov/37175032/

39. Xian S, Chen K, Cheng Y. Improved seagull optimization algorithm of partition and XGBoost of prediction for fuzzy time series forecasting of COVID-19 daily confirmed. Advances in Engineering Software. 2022 Nov 1;173.

40. Prakash S, Pathak P, Jalal AS. Predicting COVID-19 Fourth Wave Incidence in India Using Machine Learning Algorithms and SEIR Model. 9th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, UPCON 2022. 2022;

41. Melin P, Sánchez D, Monica JC, Castillo O. Optimization using the firefly algorithm of ensemble neural networks with type-2 fuzzy integration for COVID-19 time series prediction. Soft comput. 2023 Mar 1;27(6):3245–82.

42. Bhattacharyya A, Chakraborty T, Rai SN. Stochastic forecasting of COVID-19 daily new cases across countries with a novel hybrid time series model. Nonlinear Dyn. 2022 Feb 1;107(3):3025–40.

43. Jin W, Dong S, Yu C, Luo Q. A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning. Comput Biol Med. 2022 Jul 1;146.

44. Qu Z, Li Y, Jiang X, Niu C. An innovative ensemble model based on multiple neural networks and a novel heuristic optimization algorithm for COVID-19 forecasting. Expert Syst Appl. 2023 Feb 1;212.

45. Braga M de B, Fernandes R da S, Souza GN de, Rocha JEC da, Dolácio CJF, Tavares I da S, et al. Artificial neural networks for short-term forecasting of cases, deaths, and hospital beds occupancy in the COVID-19 pandemic at the Brazilian Amazon. PLoS One. 2021;16(3):e0248161.

46. Shastri S, Singh K, Deswal M, Sachin Kumar •, Mansotra V. CoBiD-net: a tailored deep learning ensemble model for time series forecasting of covid-19. [cited 2024 Sep 12]; Available from: https://doi.org/10.1007/s41324-021-00408-3

47. Cramer EY, Ray EL, Lopez VK, Bracher J, Brennen A, Castro Rivadeneira AJ, et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. Proc Natl Acad Sci U S A [Internet]. 2022 Apr 12 [cited 2024 Sep 12];119(15). Available from: https://pubmed.ncbi.nlm.nih. gov/35394862/

48. Ke W, Lu Y. Ensemble Prediction Method Based on Decomposition–Reconstitution–Integration for COVID-19 Outbreak Prediction. Mathematics 2024, Vol 12, Page 493 [Internet]. 2024 Feb 4 [cited 2024 Sep 12];12(3):493. Available from: https://www.mdpi.com/2227-7390/12/3/493/htm

49. Roster K, Connaughton C, Rodrigues FA. Forecasting new diseases in low-data settings using transfer learning. Chaos Solitons Fractals. 2022 Aug 1;161:112306.

50. Chakraborty D, Goswami D, Ghosh A, Chan J, Ghosh S. Learning from Others: A Data Driven Transfer Learning based Daily New COVID-19 Case Prediction in India using an Ensemble of LSTM-RNNs. ACM International Conference Proceeding Series [Internet]. 2021 Jun 29 [cited 2024 Sep 12]; Available from: https://dl.acm.org/doi/10.1145/3468784.3470769

51. Panagopoulos G, Nikolentzos G, Vazirgiannis M. Transfer Graph Neural Networks for Pandemic Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence [Internet]. 2021 May 18 [cited 2024 Sep 12];35(6):4838–45. Available from: https://ojs.aaai.org/index.php/AAAI/article/view/16616

52. Li Y, Jia W, Wang J, Guo J, Liu Q, Li X, et al. ALeRT-COVID: Attentive Lockdown-awaRe Transfer Learning for Predicting COVID-19 Pandemics in Different Countries. J Healthc Inform Res [Internet]. 2021 Mar 1 [cited 2024 Sep 12];5(1):98–113. Available from: https://link.springer.com/article/10.1007/s41666-020-00088-y

53. Gautam Y. Transfer Learning for COVID-19 cases and deaths forecast using LSTM network. ISA Trans. 2022 May 1;124:41–56.

FINANCING
No financing.

CONFLICT OF INTEREST
None.

AUTHORSHIP CONTRIBUTION
Conceptualization: Satya Prakash.
Data curation: Satya Prakash.
Formal analysis: Satya Prakash.
Research: Satya Prakash.
Methodology: Satya Prakash.
Supervision: Pooja Pathak, Anand Singh Jalal.
Validation: Pooja Pathak, Anand Singh Jalal.
Drafting - original draft: Satya Prakash.