Time Series Analysis with Systematic Survey on Covid-19 Based Predictive Studies During Pandemic Period using Enhanced Machine Learning Techniques

https://doi.org/10.3991/ijoe.v19i07.39089

K. Rajeswari, Sushma Vispute(✉), Amulya Maitre, Reena Kharat, Amruta Aher, N. Vivekanandan, Renu Kachhoria, Swati Jaiswal
Pimpri Chinchwad College of Engineering, Pune, India
sushma.vispute@pccoepune.org

Abstract—Coronavirus 2 virus is responsible for the spread of the infectious disease COVID-19 (also known as Coronavirus disease). People around the globe who got infected with the virus experienced a respiratory illness that could become as serious as leading someone to lose their life. However, the upside of the pandemic is that it has led to numerous types of research and explorations, majorly in the medical science field. Since a systematic survey of previous research activities and bibliometric analysis gives a brief idea about such contributions and acts as a reference to future research, this study aims to cover the research related to COVID-19 in the computer technology domain. It is limited to the works accepted and accessible with the keywords – Covid-19, prediction, and pandemic, in the Scopus search engine to justify the scope of this survey. Further, the paper highlights a few prior works used for predictive analysis and presents a quantitative angle on their algorithms. Earlier works showcase Time Series Analysis using ARIMA/SARIMA models for predicting the vaccination rates, and Extreme Gradient Boosting (XGBoost), Xtremely Boosted Network (XBNet) Regression, and Recurrent Neural Network (RNN) for Confirmed, Cured, and Death cases. Amongst the algorithms used in the latter use case, XBNet regression performed better than XGBoost regressor.

Keywords—Covid-19, pandemic, prediction, time series analysis, predictive machine learning, regressor, vaccination rate

1 Introduction

Coronavirus is a type of SARS virus. SARS is an abbreviation for the severe acute respiratory syndrome. The world has witnessed the deadly effects of this virus, thus leading to the declaration of the pandemic. The virus affected different people with varying severity. Patients with mild to moderate illness, with common symptoms like cough, fever, loss of taste/smell, and tiredness, recovered while quarantined without being hospitalized. However, patients with severe symptoms like chest pain, shortness of breath or difficulty breathing or loss of speech/mobility, or confusion had to seek immediate medical attention [49]. Not only has it affected people's well-being, but also impacted various aspects of society worldwide, thus indicating its stringency. Countries worldwide undertook different measures to overcome the aftermath of the pandemic. Basic practices assumed wearing a mask, regular sanitization, and social distancing to minimize the spread.

Since the average time for virus symptoms to surface after infection is 5–6 days, or as long as 14 days, it becomes challenging to track its severity. Therefore, this became a base use case where predictive analysis could play a vital role in understanding the situation better, and evidently, research rapidly grew in this aspect.

The contributions in this paper are outlined as follows:

- Present a bibliometric review on research from Scopus associated with predictive studies surrounding COVID-19 in the computer technology domain with illustrative statistics under various categories.
- Highlight a few prior works used for predictive analysis with a quantitative angle on their algorithms and showcase their usage in certain predictive use cases identified under Covid-19 circumstances.
- Layout probable future scope in this domain.

The paper constitutes various sections. Section 1 discusses an overview of Covid-19, its effects on human health and worldwide, precautionary measures, and the objectives of this paper. The literature review and the bibliometric study are presented in Section 2, followed by Section 3, which expresses the use of Time Series Analysis using various enhanced Machine Learning techniques in prior work for different use cases. The study concludes with an overview of topics covered throughout and mentions potential use cases that could be addressed in the future using predictive analytics in this domain.

1.1 Predictive studies

A mathematical process that aims to predict future possibilities or outcomes by analysing patterns potentially to forecast future results is known as predictive modelling or predictive analytics. The most basic question to be answered while applying predictive modelling is that based on past behaviour, what could be its future consequence. For such predictive analytics, time series analysis is used, where the data points are illustrated at successive time intervals. Using this technique, one can predict future events by analysing past trends and further extrapolating them.

Figure 1 shows a general workflow of the predictive model for Time Series Analysis. Considering the situation of Covid-19, the usual approach would be to gather a dataset relating to one of the use cases for predictive analysis, like forecasting the number of Covid-19 cases in the upcoming period. After the application of certain pre-processing on the data, typically, an exploratory analysis with data visualization is done to understand the data in a better way. With the comprehension of data, the relevant and most significant information is captured in feature engineering and then utilized to train the model, in the case of CNN-based models. Further, after model evaluation, the model is set to predict the future.

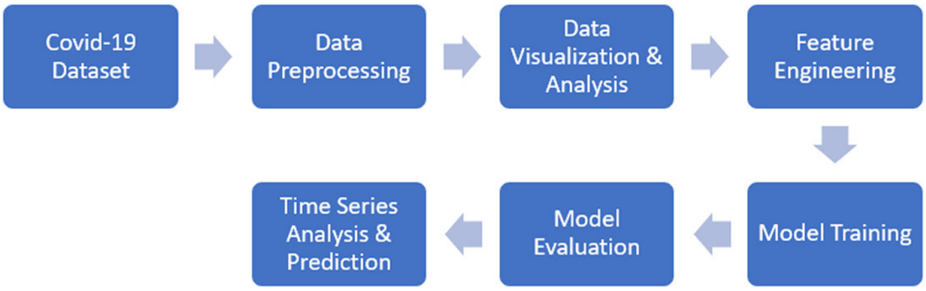

Fig. 1. Predictive model for time series analysis

2 Literature survey

2.1 Review of research activities

The pandemic has led various researchers worldwide to perform research based on different aspects and outcomes of the Covid-19 pandemic. The study focuses on showcasing the predictive analysis on Covid-19 during the pandemic (2020–2022) as shown in Table1. The global COVID-19 outbreak has resulted in a catastrophic loss of life and poses an unprecedented threat to food systems, global health, and the workplace [10–11]. Tens of millions of people experience abysmal poverty, and the number of undernourished people, which is presently projected to be close to 690 million, might climb by as many as 132 million by the end of every year. Rajaraman et. al expressed that that the epidemic has wreaked havoc on the economic structure [18]. Over half of the world's 3.3 billion employees' jobs are under peril.

There are numerous studies [2, 4, 6, 8, 15, 21–25, 27, 33, 37, 44–46, 50–51, 53] that have presented predictive models for various problems laid by the Covid-19 pandemic. Some of them include predicting the severity of the infection, understanding the patients' recovery rate, forecasting the Covid-19 spread worldwide, respiratory decompensation, etc. Deep learning techniques have become more popular in recent years, radically changing the landscape of several academic disciplines. Many researchers have opted for image processing-based andmachine/deep learning-based approaches [3, 5, 7, 11, 18, 32, 34, 36, 52] to obtain results for prediction and classification problemssincethesetechniques may minimise baseless and adverse errors in the detection and diagnosis of Covid-19, providing a new potential to serve patients with rapid, inexpensive, and safe diagnostic services. Some research is done on molecular docking [1, 26] and a thorough risk assessment of Covid-19 [24]. It is also interesting to know that there are some views on how emerging technologies like the Internet of Things (IoT) [4, 10, 19] can aid in addressing Covid-19 related issues. Collectively, a lot of review studies have given a clear picture of Covid-19 and how it can be tackled with the help of technology [9–10, 12–14, 16, 20, 28–31, 35, 38].

Table 1. Overview of research areas

Ref. Papers	Area of Research
[2, 4, 6, 8, 15, 21–25, 27, 33, 37, 44–46, 50–51, 53]	Machine learning-based approach for prediction/classification problems
[3, 5, 7, 11, 18, 32, 34, 36, 52]	Different deep learning/image processing methods for dealing with issues related to Covid-19
[1, 17, 26]	Molecular Docking
[24]	Risk assessment of Covid-19
[4, 9–10, 12–14, 16, 20, 28–31, 35, 38]	Reviews on emerging technologies

2.2 Scopus bibliography analysis

This study is limited to the research work accepted by Scopus [39] and thus, giving the scope of availing the Scopus search engine.

Preliminary data collection. A bibliometric review gives brief insights about the recent research work done in the considered subject area. In this survey, the attention is on predictive analyses based on Covid-19 done during the pandemic. The Scopus search engine helped capture information using some momentous keywords listed in the upcoming sub-section. The publications are filtered using these keywords to obtain desired results. Further, the paper presents a graphical analysis based on various parameters.

Momentous keywords. "Covid-19" and "Prediction" are the primary keywords required to search. The secondary keyword is "Pandemic", with the limit to the subject area of "Computer Science". Table 2 shows this information.

Search Query is formulated like:

(TITLE-ABS-KEY (Covid-19 AND Prediction)) AND (Pandemic) AND
(LIMIT-TO (SUBJAREA, "COMP"))

Table 2. Designed search scheme for keywords

Essential Keywords (AND)	**"Covid-19" and "Prediction"**
Subordinate Keyword (AND)	**"Pandemic"**
Limit To (AND)	**SUBAREA, "COMP"**

Search results. The initial search using primary keywords filtered out 4,367 publications. Figure 2 shows the graph of the year-wise publications generated by this query between 2020 to 2022, and Table 3 tabulates the same.

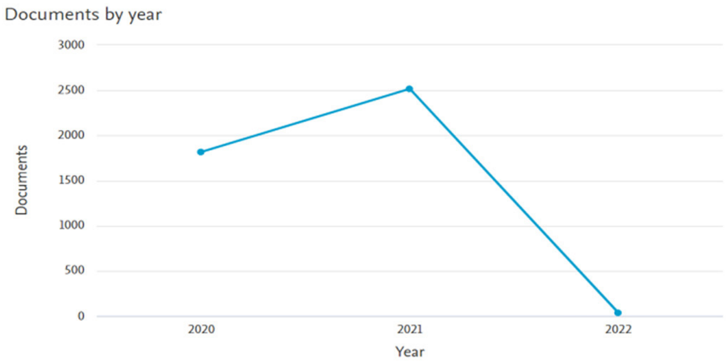

Fig. 2. No. of publications over 2020–2022 for the initial query

Source: http://www.scopus.com (September 2021).

Table 3. No. of publications year-wise for the initial query

Year of Publication	Number of Publications
2022	35
2021	2516
2020	1816
TOTAL	**4367**

With the final query, 919 publications are filtered for 2020–2022, as shown in Figure 3 and Table 4.

Table 4. No. of publications year-wise for the final query

Year of Publication	Number of Publications
2022	25
2021	610
2020	284
TOTAL	**919**

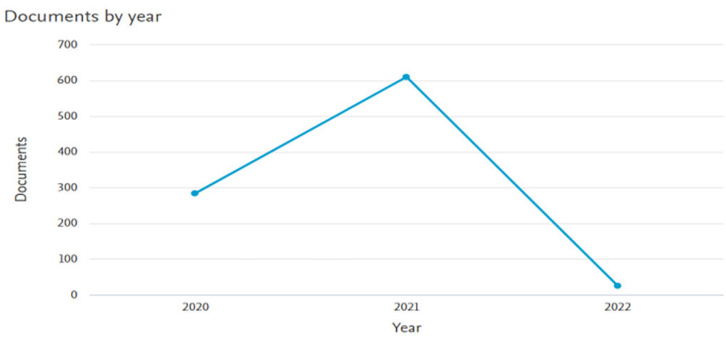

Fig. 3. No. of publications over 2020–2022 for the final query

Source: http://www.scopus.com (September 2021).

Geographical analysis. There are many publications submitted on this area of study from various countries over the world. The countries leading in this aspect are India (251), the United States (155), China (112), followed by Saudi Arabia (67), and the United Kingdom (54). Figure 4 shows the top 10 countries that have the most publications.

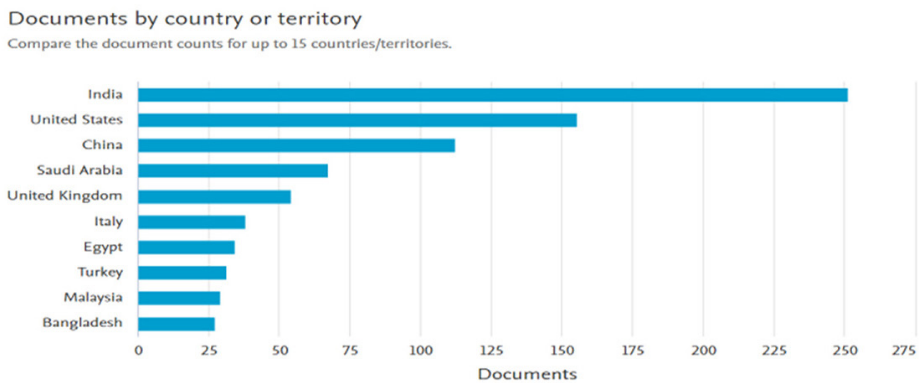

Fig. 4. Publications by countries survey

Source: http://www.scopus.com (September 2021).

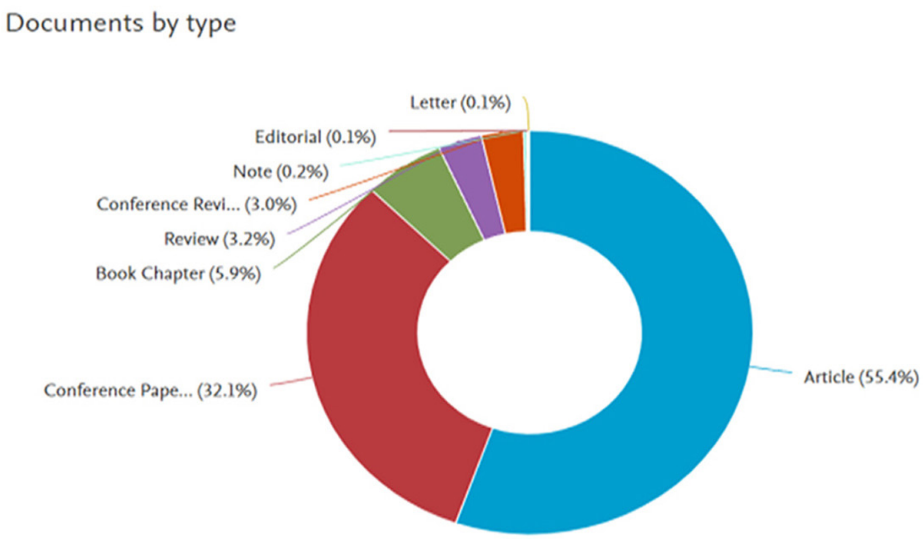

Fig. 5. Publications by document type survey

Source: http://www.scopus.com (September 2021).

Document type survey. Figure 5 demonstrates the percentage of research published in various documents like articles, conference papers, book chapters, reviews, conference reviews, etc. Most of the work is published as articles, with a publication's percentage of 55.4%, followed by Conference Papers at 32.1%. Notes, editorials, and letters take up 0.1% of the whole.

Subject area-wise analysis. Although the search is filtered to focus on "Computer Science, yet there are related subject areas like Mathematics, Engineering, Decision Sciences, and so on to support it. The subject-wise analysis reveals the connection among the subject areas. Figure 6 depicts the number of research papers published based on these subject areas.

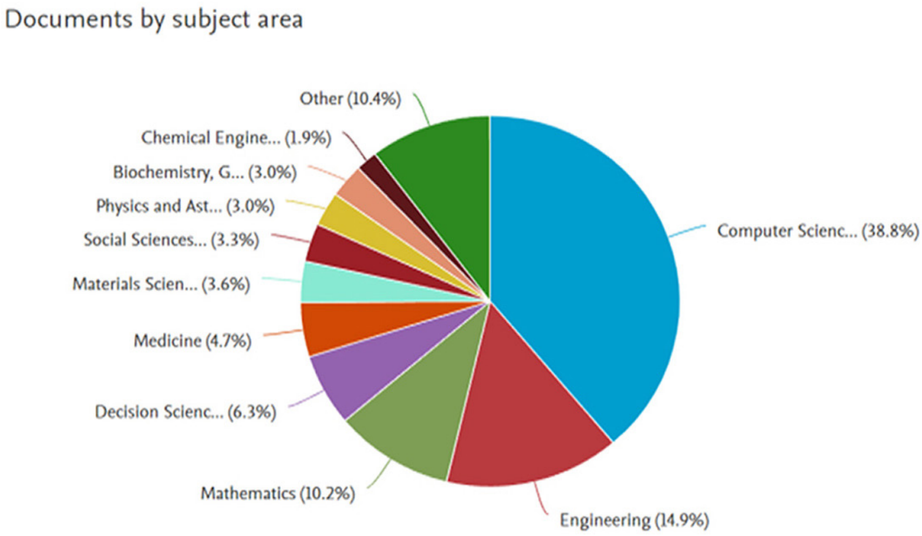

Fig. 6. Publications by subject area survey

Source: http://www.scopus.com (September 2021).

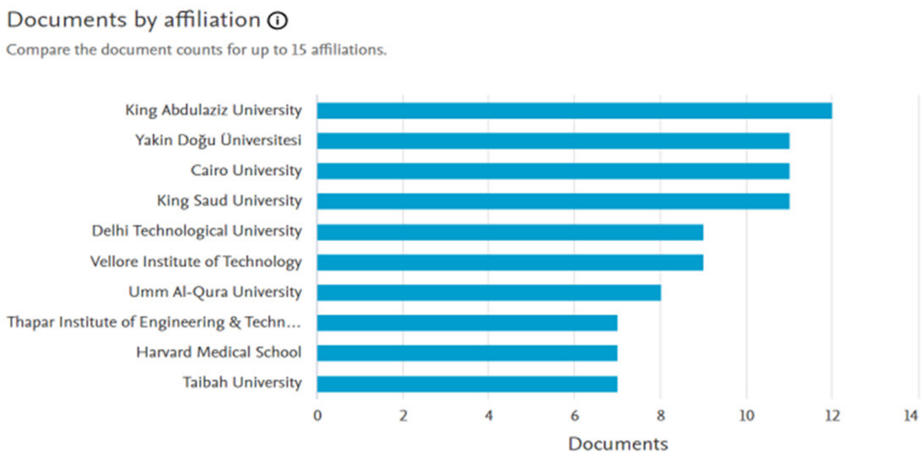

Fig. 7. Review of publications in view of affiliations

Source: http://www.scopus.com (September 2021).

Affiliations analysis. There are a lot of benefactions from various universities worldwide. It is admirable that the researchers have continued their search for the unknown even in these difficult times of pandemic. The King Abdulaziz University has provided a generous contribution among the top 10 universities, as noticed in Figure 7.

Sources analysis. The overall submissions in IEEE Access are more than the other journals/conferences when analysed based on their source of publication. Figure 8 shows the year-wise analysis, and Table 5 depicts the total submissions for the top 5 publishers.

Table 5. No. of submissions in top 5 sources

Sources	Number of Submissions
IEEE Access	31
Computers Materials and Continua	27
Computers In Biology and Medicine	24
Lecture Notes in Networks and Systems	24
ACM International Conference Proceeding Series	22

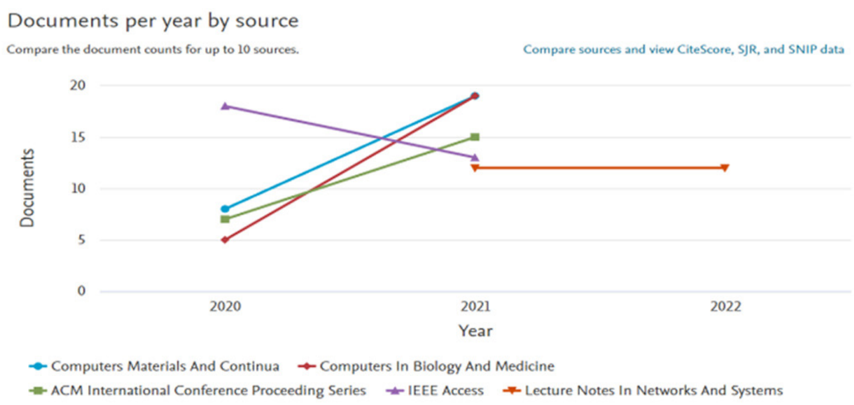

Fig. 8. Documents by sources survey

Source: http://www.scopus.com (September 2021).

Author – wise publications analysis. Figure 9 shows the top 10 authors who contributed the most to the data accessed from the Scopus search engine. Author Al-Turjman, F. has seven documents to his name on Covid-19 based predictive studies done during 2020–2022, which is the largest number of sole contributions.

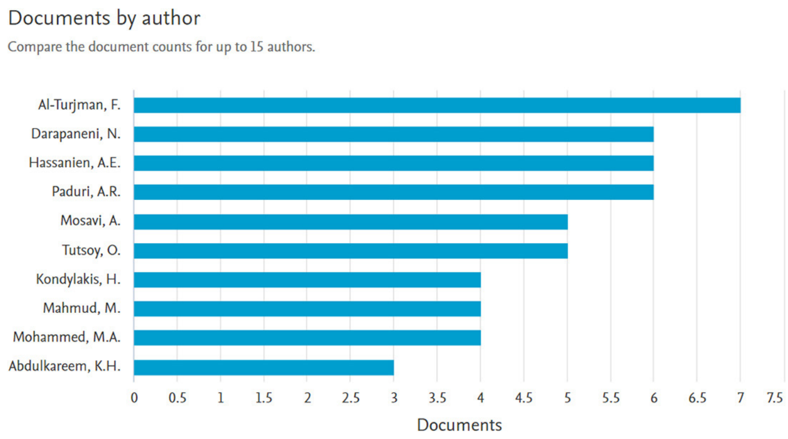

Fig. 9. Review in light of author-wise publications

Source: http://www.scopus.com (September 2021).

Citation analysis. The citation refers to other researchers' work published in a book, journal, article, or research paper. The citation analysis shows the frequency of mentions of a research document. Citation is done as per its relevance and signifies its influence on the researchers' study.

Figure 10 shows year-wise citations for various research documents. The citation count in 2020, 2021, and 2022 is 477, 2566, and 47, respectively. Figure 11 illustrates the information of the top 10 publications in increasing order of their citation count.

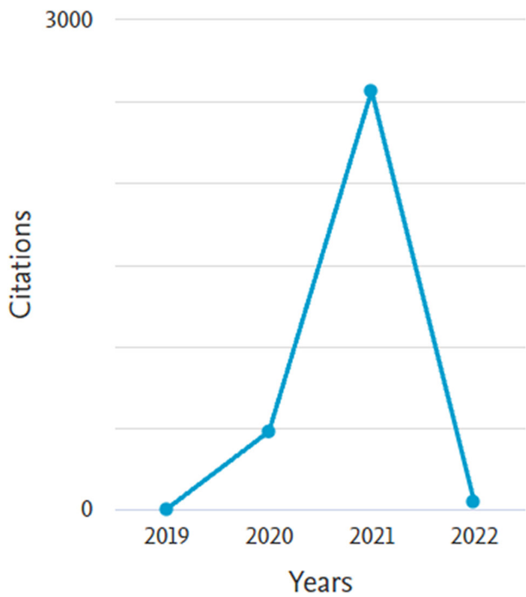

Fig. 10. Review of papers being cited per annum

Source: http://www.scopus.com (September 2021).

	Documents	Citations	<2019	2019	2020	2021	>2021	Subtotal	Total
		Total	0	0	477	2566	47	3090	3090
1	Rapid Identification of Potential Inhibitors of SARS-CoV-2 M...	2020			91	101		192	192
2	Towards an artificial intelligence framework for data-driven...	2020			60	82	2	144	144
3	Deep Learning for Classification and Localization of COVID-1...	2020			32	83		115	115
4	Predicting the growth and trend of COVID-19 pandemic using m...	2020			32	76	1	109	109
5	Deep-COVID: Predicting COVID-19 from chest X-ray images usin...	2020			9	97	2	108	108
6	COVID-19 Future Forecasting Using Supervised Machine Learnin...	2020			24	67	2	93	93
7	CovXNet: A multi-dilation convolutional neural network for a...	2020			10	74	2	86	86
8	COVID-19 outbreak prediction with machine learning	2020			30	47	3	80	80
9	Role of biological Data Mining and Machine Learning Techniqu...	2020			13	51	1	65	65
10	Predicting COVID-19 in China Using Hybrid AI Model	2020			23	33	1	57	57

Fig. 11. Citation analysis – top 10 publications

Source: http://www.scopus.com (September 2021).

2.3 Network analysis

Network analysis depicts the visualization of the datagraphically. The nodes are the data points (variables) displayed in the graph plot, and the edges exhibit the relationship between the nodes. There are various tools to create the network graphs for analysis like ScoNetV, Cuttlefish, Sciencescape, Gephi, and VOS viewer. The tool used in this survey is VOS viewer [40]. Figures 12, 13 and 14 are some network analyses visualizations created from extracted data of Scopus using different parameters. These visualizations are laid out using several layout algorithms and manual adjustments.

Network analysis on keywords. Figure 12 depicts the network-based analysis for the Cluster of Co-occurrence of Author keywords. Here, the assumed minimum number of occurrences of a keyword is 5. 96 keywords out of 2354 meet the threshold. The representation of the cluster contains 63 items and 8 clusters that depict the co-occurrences of authors' keywords. As per observation, the keywords having the maximum number of occurrences are Covid-19, deep learning, machine learning, and artificial intelligence.

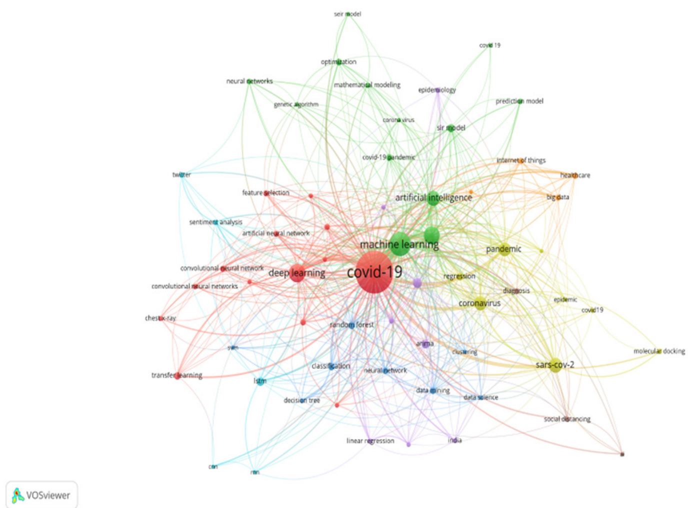

Fig. 12. Clustered representation of concurrence of author keyword

Source: http://www.scopus.com (September 2021).

Country-wise network analysis. Figure 13 depicts the representation in view of the various countries that contributed immensely to the said research area. The supposed minimum number of documents per country is 5. 46 countries out of 136 satisfied the threshold value. Representation of the cluster consists of 46 items and 7 clusters that represent the country-wise contribution to the research. As per the generated network analysis graph, most contributions are from India, followed by the United States, China, and Saudi Arabia.

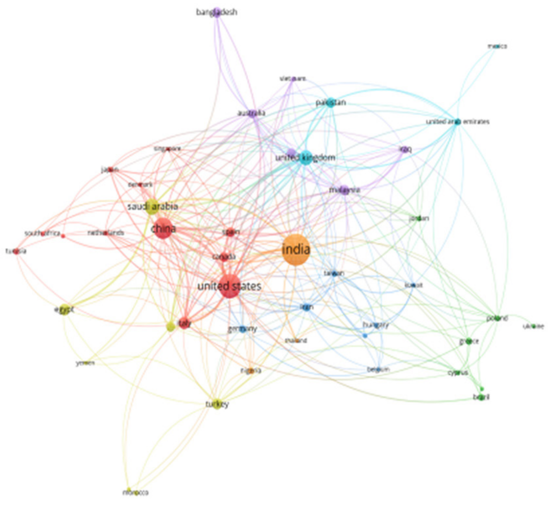

Fig. 13. Network representation diagram in view of countries who contributed to this domain
Source: http://www.scopus.com (September 2021).

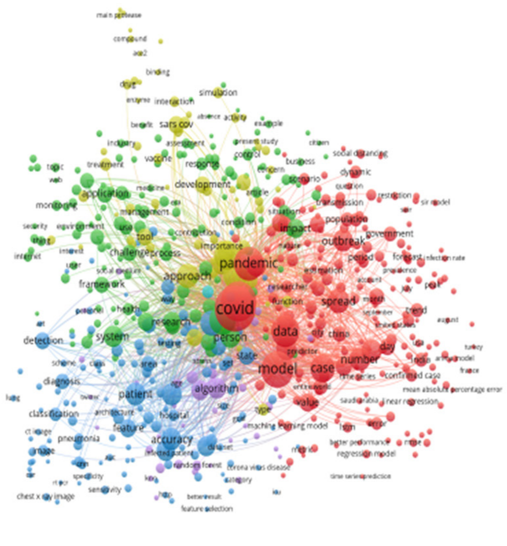

Fig. 14. Network of extracted data from the title and abstract field
Source: http://www.scopus.com (September 2021).

Figure 14 shows the representation in view of the data extracted from the title and abstract. The keywords covid, pandemic, model, data, case, number, the spread is the most prominent in the network and aligned to the momentous keywords.

3 Time series analysis using enhanced machine learning techniques

Various services and applications in real-world settings have used machine learning techniques. Applications in disease and healthcare analytics, such as coronavirus disease identification and predictive analytics, are among them. Via the above-mentioned extensive literature survey, a few notable and commonly used algorithms to perform a time series analysis are elaborated with their usages on different Covid-19 datasets in this section.

3.1 Recurrent Neural Network (RNN)

One of the types of Artificial Neural Networks (ANN) is Recurrent Neural Networks (RNN). Although RNN has a broad scope of usage, here, we specify its ability to analyse sequential or time-series data [41]. RNN relies on the previous elements within the sequence to obtain the output, while conventional Deep Neural Networks (DNN) expect that information sources are autonomous of the resultant. While approaching occasions would likewise help with deciding the result of a given arrangement, unidirectional recurrent neural networks can't represent these occasions in their anticipation. RNN is considered to have exceptional potential in time-series forecasting due to its ability to save extensive past information in its internal state. However, it has the restrictions of vanishing, and gradient-exploding issues, which leads to an ineffective practiceor prolonged practice period. Hence, RNN has different variants amongst, which are the Long short-term memory (LSTM), Bidirectional Recurrent Neural networks (BRNN), and Gated recurrent units (GRUs). [44, 45, 50, 53] have used RNN/LSTM to experiment with various Covid-19 use cases.

3.2 ARIMA

An ARIMA, i.e., Auto-Regressive Integrated Moving Average, model demonstrates a time series based on its previous data, i.e., the model's equation is constructed with its own lags and lagged prediction errors. This model consists of three terms, *p* (order of AR term), *d*, and *q* (order of MA term). The term d makes the time series stationary. The term d makes the time series stationary. It determines the number of differences required [42, 44, 46].

In a pure AR (Auto-Regressive) model, Y_t, depends on the lags, giving us the first half of the equation [44].

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_1 \tag{1}$$

Here, the 'lags of Y_t' are a function of Y_t. For instance, the first lag is Y_{t-1} in the series, with β_1 as its coefficient, which the model calculates. α is the intercept term that the model also intercepts.

Similarly, the equation of a pure MA (Moving Average) model gives the second half of the final equation, where Y_t is solely dependent on the lag in forecast errors, as shown below. Here, the error terms from the autoregressive models are the errors of the individual lags'.

$$Y_t = \alpha + \epsilon_1 + \varphi_1 \epsilon_{t-1} + \varphi_2 \epsilon_{t-2} + \cdots + \varphi_q \epsilon_{t-q} \tag{2}$$

The errors ϵ_t and ϵ_{t-1} can be expressed as follows:

$$\begin{aligned} Y_t &= \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_0 Y_0 + \epsilon_1 \\ Y_{t-1} &= \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + \cdots + \beta_0 Y_0 + \epsilon_{t-1} \end{aligned} \tag{3}$$

So, the equation for the final ARIMA is:

$$\begin{aligned} Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_1 \\ + \varphi_1 \epsilon_{t-1} + \varphi_2 \epsilon_{t-2} + \cdots + \varphi_q \epsilon_{t-q} \end{aligned} \tag{4}$$

Experimentsperformed by [46, 51, 53] to predict the vaccination ratesand infected/death casesfor variouscountries using ARIMA and SARIMA models.

3.3 SARIMA

The SARIMA, i.e., Seasonal ARIMA, and SARIMAX, i.e., Seasonal ARIMA with "eXogenous" factors models, are the extensions to the ARIMA model. The hardship with the basic ARIMA model is that the seasonality factor isn't taken into consideration. A seasonal differencing model, such as SARIMA or Seasonal ARIMA is used when a time series exhibits seasonal patterns. Differencing in the seasonal model is analogous to regular differencing, where a value is subtracted from the prior season instead of deducting the consecutive terms. Hence, SARIMA is defined as SARIMA (p, d, q) x (P, D, Q) [44]. Here, the terms P, D are the order of SAR and Q is the SMA seasonal differencing terms' order. The frequency of the time series is denoted by 'x'. Set D = 1 if the model has apparent seasonal patterns for a given frequency 'x'.

3.4 Regression analysis

One of the popular statistical techniques is Regression analysis. This technique helps establish an association between two variables. In Linear Regression, two factors are connected by an equation where the exponent of the two factors is 1. When plotted as a chart, a linear relationship addresses a straight line numerically. In a non-linear relationship, a bend is framed as the exponent of any factor that is not equivalent to one. The accompanying, condition 5, is the overall numerical condition for linear regression.

$$P = aq + b \tag{5}$$

In the equation (5), P is the dependent (response) variable, q is the independent (predictor) variable, and constants, a and b are known as coefficients. Equation (6) represents the formula for the loss function of the linear regression model.

$$L_s = \frac{1}{2}\sum_{i=0}^{n}(b + aqi - pi)^2 \tag{6}$$

Extreme gradient boosting (XGBoost) and xtremely boosted network (XBNet) regressor. XGBoost is an efficient and effective gradient boosting algorithm implementation. The purpose of regression predictive modelling issues is to forecast a numerical value, such as a dollar amount or a height. XGBoost can be utilised immediately for regression predictive modelling [47].

XBNet is an open-source project built with PyTorch that aims to create a robust architecture by combining tree-based models and neural networks. Its advantages include high performance, accelerated training, and faster inference speed. Its rapid prototyping capabilities make this simple to implement [48].

3.5 Application of various algorithms for COVID-19 data analysis

A lot of problems had been introduced during the pandemic period which led an opportunity for researchers from around the world to ponder upon them and come up with effective solutions. In this section, few of these use cases have been discussed. K. Rajeswari et. al [44] have provided the comparative findings of variousdeep learning and statistical methods for predicting Covid-19 cases. Traditionally, the long-term performance of deep learning models is typically better than that of statistical models. Hence, the author has aligned their observations regarding statistical and deep learning methods with the same notion. Here, it is discussed; those statistical models tend to get outperformed by deep learning models in the long term via their results. On average, statistical methods like SARIMA and Linear Regression (LR) were better for forecasting one day and seven days, respectively, that is, they could perform better predictions for fewer days. On the other hand, the deep Learning method, LSTM, forecasted for a greater period like one month and tended to be more consistent. Table 6 shows the comparative analysis of RMSE values based on time span of one-day, seven-day, and one-month for linear regression, ARIMA, SARIMA, and LSTM on the JHU dataset. Figure 15 illustrates the graphical representation of Table 6. Figure 16 illustrates the prediction of Covid-19 cases using Long Short-Term Memory from October 25th to November 25th, 2021. These predictions indicate that the RMSE (Root Mean Square Error) for one-day is 1.95E+03, while for seven-day and one month is 1.48E+03 and 1.41E+03, respectively.

Table 6. No. of submissions in top 5 sources

Algorithm	1 Day	7 Days	1 Month
Linear Regression	2.52E+03	1.45E+03	2.76E+ 03
ARIMA	2.54E+03	2.47E+03	4.32E+03
SARIMA	1.55E+03	1.48E+03	2.46E+03
LSTM	1.95E+03	1.48E+03	1.41E+03

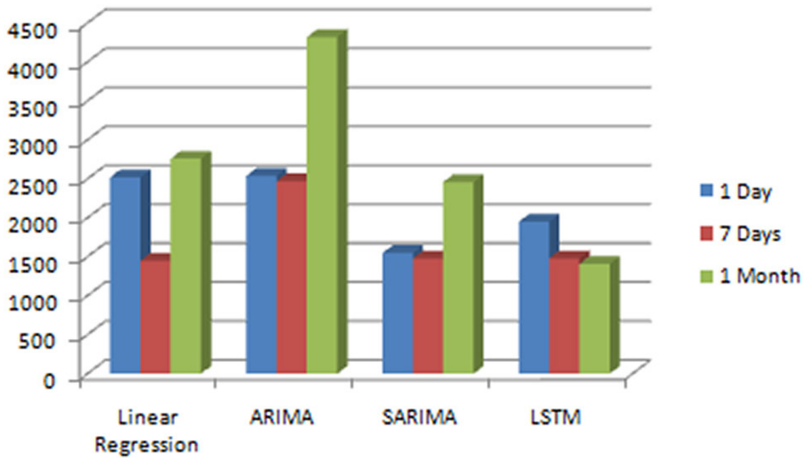

Fig. 15. Time series analysis for daily cases

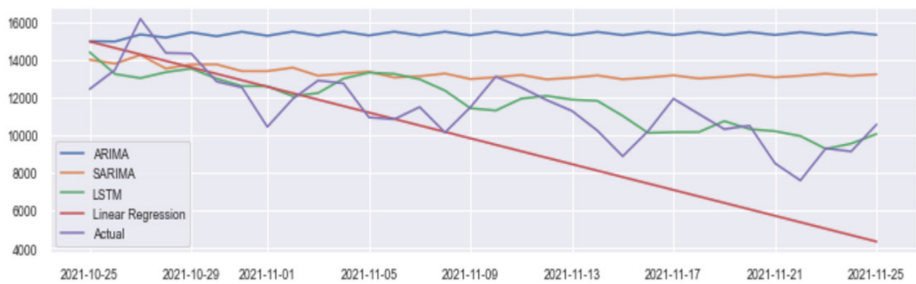

Fig. 16. Time series analysis comparison of different algorithms [44]

On similar lines, Kumar RL et. al [50] employed anIndianCOVID-19 casesdataset from Kaggle to predict the rate of COVID-19cases in future. They have proposed a deep learning-based prediction method for predicting COVID-19confirmed and death cases. The modified-LSTM model was used to forecast the same. The predicted outcomes were then optimized using DRL made based on the symptoms. Figure 17 depicts their results of confirmed and death cases estimated for next 15 days. This model is named MLSTM-DRL and is compared with Logistic Regression (LR) and LSTM using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Based on the results, the suggested MLSTM-DRL shows the lowest error rate, whereas Logistic Regression (LR) has a higher error rate. Figure 18 shows the comparison of assessment metrics values for the algorithms in consideration.

As discussed above, there are various statistical techniques to perform time series analysis like ARIMA/SARIMA. Amulya Maitre et. al [46] showcased the use of these algorithms to predict the vaccination rates in India for next 30 days based on number of people vaccinated amongst 100 people. Here, the easily accessible vaccination data is utilized from https://ourworldindata.org/coronavirus-source-data. The sample results for India are shared below in Figure 19 [46].

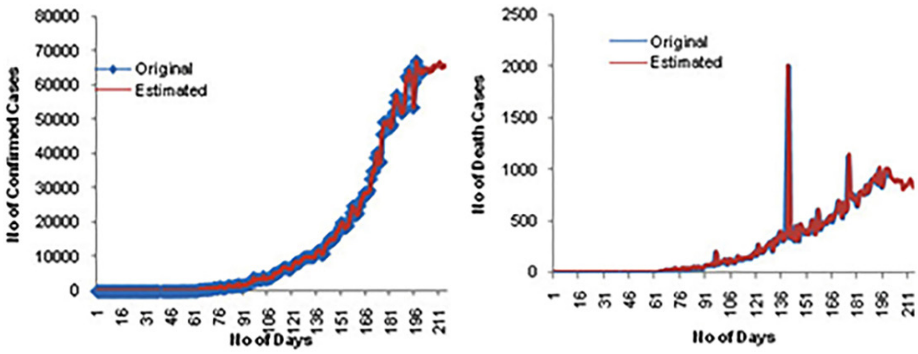

Fig. 17. Comparison of MLSTM-DRL with logistic regression and LSTM [50]

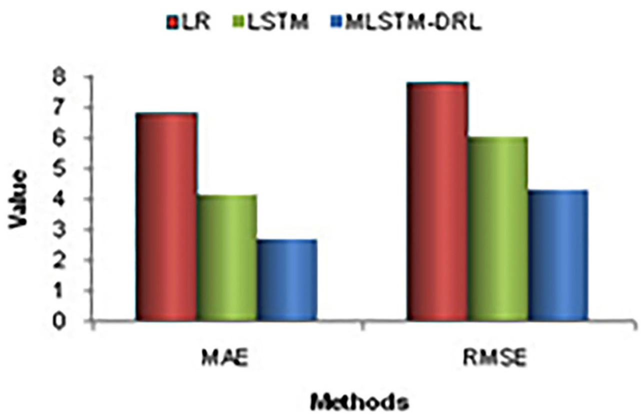

Fig. 18. Comparison of MLSTM-DRL with logistic regression and LSTM [50]

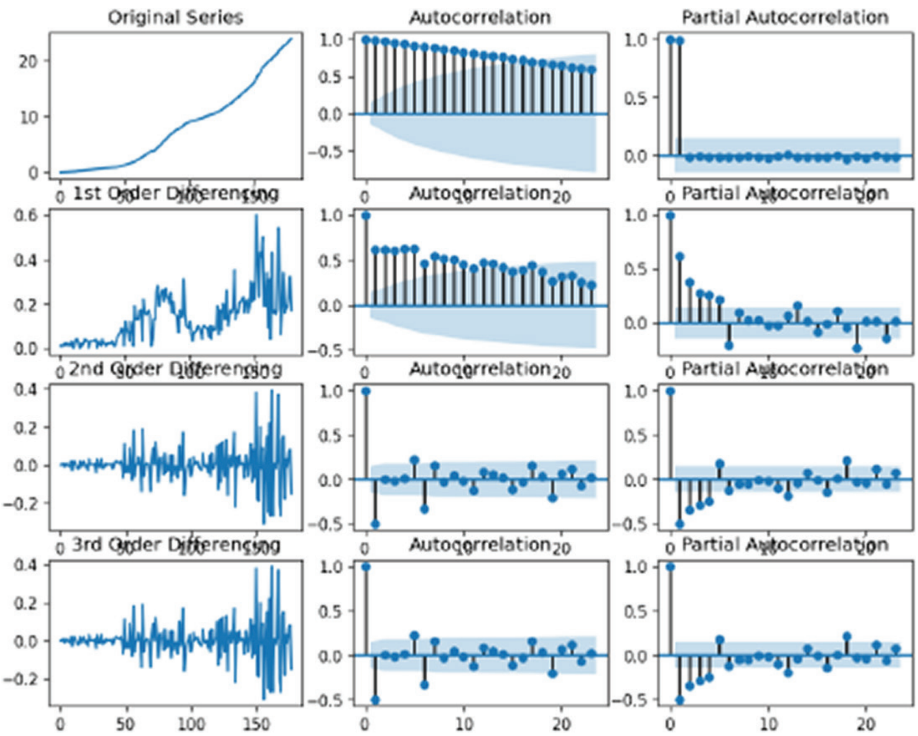

Fig. 19. a) ACF and PACF plots based on the vaccination data

```
Performing stepwise search to minimize aic
 ARIMA(1,2,1)(0,0,0)[0] intercept   : AIC=-405.821, Time=0.30 sec
 ARIMA(0,2,0)(0,0,0)[0] intercept   : AIC=-318.026, Time=0.06 sec
 ARIMA(1,2,0)(0,0,0)[0] intercept   : AIC=-368.008, Time=0.17 sec
 ARIMA(0,2,1)(0,0,0)[0] intercept   : AIC=-407.113, Time=0.14 sec
 ARIMA(0,2,0)(0,0,0)[0]             : AIC=-320.011, Time=0.07 sec
 ARIMA(0,2,2)(0,0,0)[0] intercept   : AIC=-405.951, Time=0.26 sec
 ARIMA(1,2,2)(0,0,0)[0] intercept   : AIC=-403.115, Time=0.48 sec
 ARIMA(0,2,1)(0,0,0)[0]             : AIC=-408.245, Time=0.05 sec
 ARIMA(1,2,1)(0,0,0)[0]             : AIC=-407.057, Time=0.12 sec
 ARIMA(0,2,2)(0,0,0)[0]             : AIC=-407.225, Time=0.12 sec
 ARIMA(1,2,0)(0,0,0)[0]             : AIC=-369.929, Time=0.07 sec
 ARIMA(1,2,2)(0,0,0)[0]             : AIC=-405.561, Time=0.24 sec

Best model:  ARIMA(0,2,1)(0,0,0)[0]
Total fit time: 2.106 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  180
Model:               SARIMAX(0, 2, 1)   Log Likelihood                 206.123
Date:                Wed, 21 Jul 2021   AIC                           -408.245
Time:                        15:49:48   BIC                           -401.882
Sample:                             0   HQIC                          -405.665
                                - 180
Covariance Type:                  opg
```

Fig. 19. b) Results of the models after application on the dataset

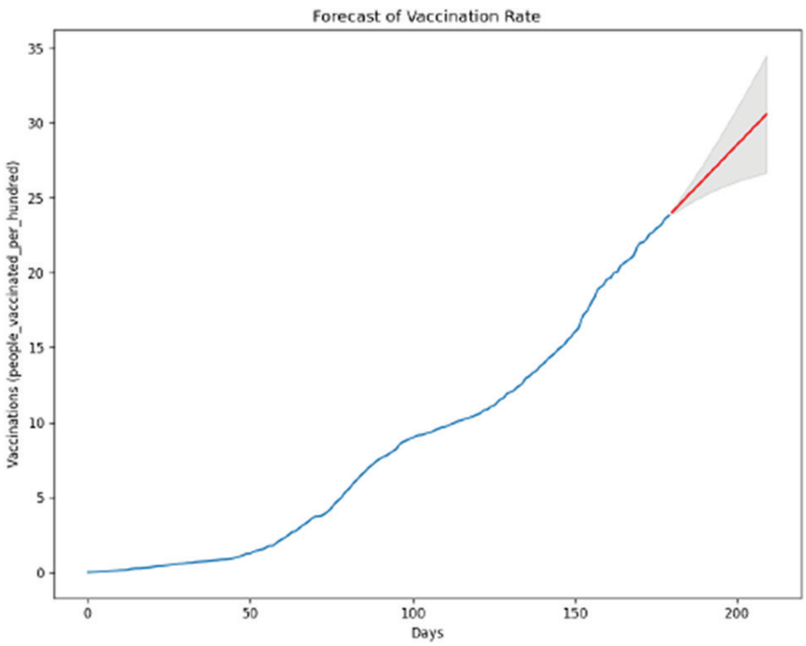

Fig. 19. c) Forecast of vaccination rate for India

Another application of XGBoost Regressor and XBNetRegressor models was identified on Kaggle. Figure 20 (a–d) shows Covid-19 cases analysis of COVID-19 dataset of India taken from Kaggle website for "Confirmed", "Cured" and "Deaths" cases respectively [43]. Figure 21 shows F-score generated using XGBoost model for Cured and Death cases and Figure 22 shows the XBNetRegressor Accuracy and loss for the COVID-19 India dataset from the Kaggle site [43].

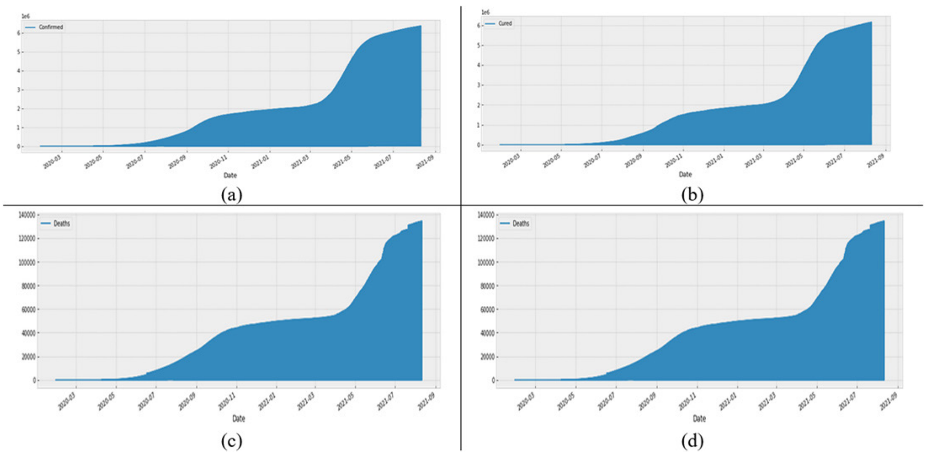

Fig. 20. Covid-19 cases analysis of COVID-19 dataset of India taken from Kaggle

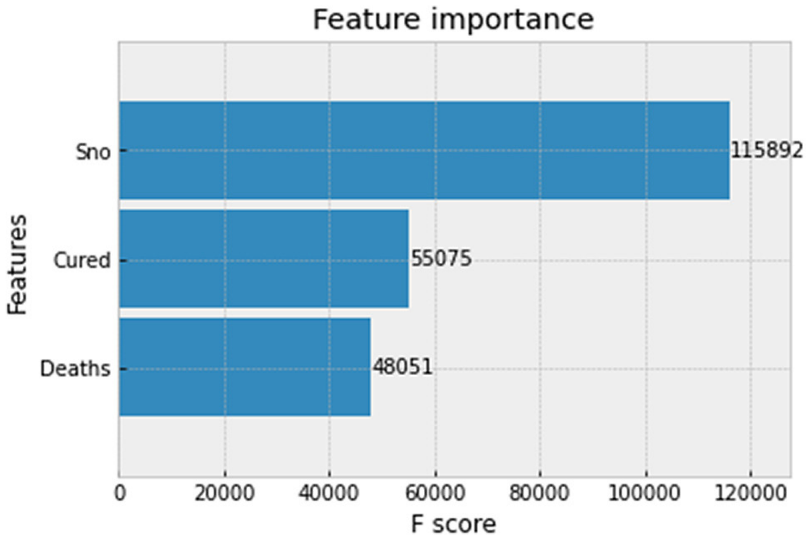

Fig. 21. F-score generated using XGBoost model for cured and death cases [43]

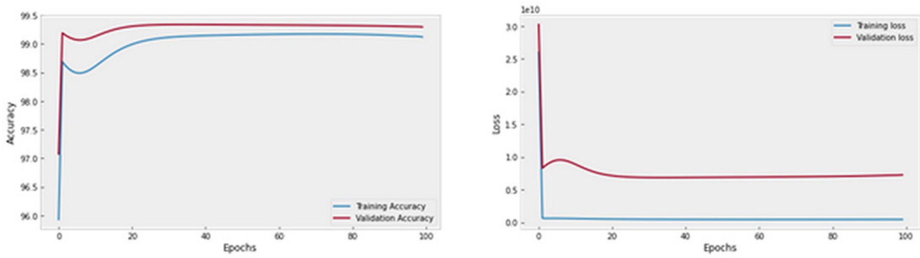

Fig. 22. XBNetRegressor accuracy and loss for Covid-19 India dataset [43]

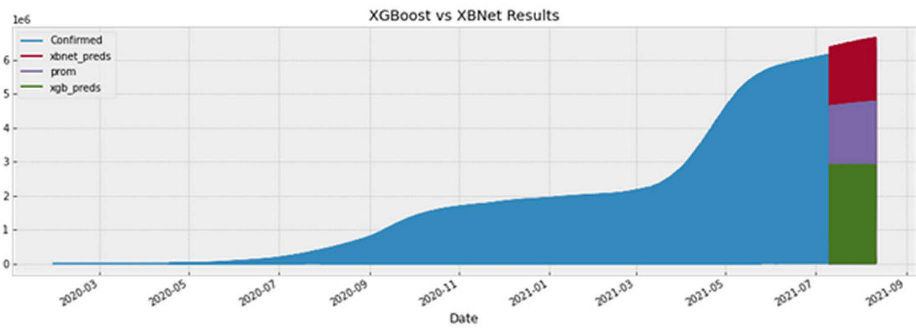

Fig. 23. Accuracy comparison between XGBoost and XBNetRegressor
for the Covid-19 India dataset [43]

Figure 23 shows an accuracy comparison between XGBoost and XBNetRegressor for the covid-19 India dataset [43]. The results show that XBNet regressor outperforms the XGBoost Regressor.

There is much more research work that similarly shows predictive analysis on different Covid-19 datasets to address various use cases. The significant observation is that research is not just limited to statistical algorithms but has also extended to the usage of various deep learning approaches to consume their abilities. This deduces the fact that machine learning and deep learning methods are as applicable in the medical domain as in any other field of research.

4 Conclusion and future scope

As a result of this systematic survey and bibliometric study, we conclude that the research on Covid-19 drastically increased through the pandemic where India is leading in the Covid-19 research for predictive analytics. During the pandemic era, extensive research was evolving around medical science, which benefitted humankind in this disaster. However, to understand the catastrophic situation, some studies shed light on various predictive analyses possible in that scenario. In the same context, several topics are either unturned or least researched. This survey highlights that the areas like the internet of things, epidemiology, transfer learning, big data, etc., are identified as least studied. The extension of such research work leads to endless possibilities for exploration. This bibliometric study illustrates statistics under various categories, which could be helpful for future research in this aspect. Furthermore, few prior works used for predictive analysis with a quantitative angle on their algorithms are presented in this paper. This highlights the usage of different Time Series Analysis algorithms applied on various Covid-19 datasets identified based on use cases identified under Covid-19 circumstances.

Furthermore, as described in this study, Covid-19 impacted many work streams, hence, the probable future scope would be to determine how Covid-19 has affected work sectors like education or industries for predictive analysis. Also, studies highlighting symptoms post virus infection with its controlling measures can be carried forward.

5 References

[1] Ton, A.-T., Gentile, F., Hsing, M., Ban, F., Cherkasov, A. (2020). Rapid identification of potential inhibitors of SARS-CoV main protease by deep docking of 1.3 billion compounds. Molecular Informatics, 39 (8), art. no. 2000028. https://doi.org/10.1002/minf.202000028

[2] Jiang, X., Coffee, M., Bari, A., Wang, J., Jiang, X., Huang, J., Shi, J., Dai, J., Cai, J., Zhang, T., Wu, Z., He, G., Huang, Y. (2020). Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. Computers, Materials and Continua, 63 (1), pp. 537–551. https://doi.org/10.32604/cmc.2020.010691

[3] Roy, S., Menapace, W., Oei, S., Luijten, B., Fini, E., Saltori, C., Huijben, I., Chennakeshava, N., Mento, F., Sentelli, A., Peschiera, E., Trevisan, R., Maschietto, G., Torri, E., Inchingolo, R., Smargiassi, A., Soldati, G., Rota, P., Passerini, A., Van Sloun, R.J.G., Ricci, E., Demi, L. (2020). Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound. IEEE Transactions on Medical Imaging, 39 (8), art. no. 9093068, pp. 2676–2687. https://doi.org/10.1109/TMI.2020.2994459

[4] Tuli, S., Tuli, S., Tuli, R., Gill, S.S. (2020). Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. Internet of Things (Netherlands), 11, art. no. 100222. https://doi.org/10.1016/j.iot.2020.100222

[5] Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., JamalipourSoufi, G. (2020). Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. Medical Image Analysis, 65, art. no. 101794. https://doi.org/10.1016/j.media.2020.101794

[6] Rustam, F., Reshi, A.A., Mehmood, A., Ullah, S., On, B.-W., Aslam, W., Choi, G.S. (2020). COVID-19 future forecasting using supervised machine learning models. IEEE Access, 8, art. no. 9099302, pp. 101489–101499. https://doi.org/10.1109/ACCESS.2020.2997311

[7] Mahmud, T., Rahman, M.A., Fattah, S.A. (2020). CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. Computers in Biology and Medicine, 122, art. no. 103869. https://doi.org/10.1016/j.compbiomed.2020.103869

[8] Ardabili, S.F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A.R., Reuter, U., Rabczuk, T., Atkinson, P.M. (2020). COVID-19 outbreak prediction with machine learning. Algorithms, 13 (10), art. no. 249. https://doi.org/10.3390/a13100249

[9] Albahri, A.S., Hamid, R.A., Alwan, J., Al-qays, Z.T., Zaidan, A.A., Zaidan, B.B., Albahri, A.O.S., AlAmoodi, A.H., Khlaf, J.M., Almahdi, E.M., Thabet, E., Hadi, S.M., Mohammed, K.I., Alsalem, M.A., Al-Obaidi, J.R., Madhloom, H.T. (2020). Role of biological data mining and machine learning techniques in detecting and diagnosing the Novel Coronavirus (COVID-19): A systematic review. Journal of Medical Systems, 44 (7), art. no. 122. https://doi.org/10.1007/s10916-020-01582-x

[10] Ajaz, F., Naseem, M., Sharma, S., Shabaz, M., Dhiman, G. (2021). COVID-19: Challenges and its technological solutions using IoT. Current Medical Imaging. PMID: 33588738V. https://doi.org/10.2174/1573405617666210215143503

[11] Shriram, S., Nagaraj, B., Jaya, J., Shankar, S., Ajay, P. (2021). Deep learning-based real-time AI virtual mouse system using computer vision to avoid COVID-19 spread. Journal of healthcare engineering. https://doi.org/10.1155/2021/8133076

[12] Ajmera, K. M. (2021). Fatal case of rhabdomyolysis post-covid-19 vaccine. Infection and Drug Resistance, 14, 3929. https://doi.org/10.2147/IDR.S331362

[13] Chopra, S., Dhiman, G., Sharma, A., Shabaz, M., Shukla, P., Arora, M. (2021). Taxonomy of adaptive neuro-fuzzy inference system in modern engineering sciences. In S. H. Ahmed (Ed.), Computational Intelligence and Neuroscience, Hindawi Limited, 2021, pp. 1–14. https://doi.org/10.1155/2021/6455592

[14] Ajmera, K.M. (2021). COVID-19 and its association with new-onset diabetes. Archives of Clinical and Medical Case Reports, 5, pp. 855–861. https://doi.org/10.26502/acmcr.96550430

[15] Zheng, N., Du, S., Wang, J., Zhang, H., Cui, W., Kang, Z., Yang, T., Lou, B., Chi, Y., Long, H., Ma, M., Yuan, Q., Zhang, S., Zhang, D., Ye, F., Xin, J. (2020). Predicting COVID-19 in China using hybrid AI model. IEEE Transactions on Cybernetics, 50 (7), art. no. 9090302, pp. 2891–2904. https://doi.org/10.1109/TCYB.2020.2990162

[16] Mohamadou, Y., Halidou, A., Kapen, P.T. (2020). A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. Applied Intelligence, 50 (11), pp. 3913–3925. https://doi.org/10.1007/s10489-020-01770-9

[17] Gimeno, A., Mestres-Truyol, J., Ojeda-Montes, M.J., Macip, G., Saldivar-Espinoza, B., Cereto-Massagué, A., Pujadas, G., Garcia-Vallvé, S. (2020). Prediction of novel inhibitors of the main protease (M-pro) of SARS-CoV-2 through consensus docking and drug reposition. International Journal of Molecular Sciences, 21 (11), art. no. 3793. https://doi.org/10.3390/ijms21113793
[18] Rajaraman, S., Siegelman, J., Alderson, P.O., Folio, L.S., Folio, L.R., Antani, S.K. (2020). Iteratively pruned deep learning ensembles for COVID-19 detection in chest X-Rays. IEEE Access, 8, art. no. 9121222, pp. 115041–115050. https://doi.org/10.1109/ACCESS.2020.3003810
[19] Otoom, M., Otoum, N., Alzubaidi, M.A., Etoom, Y., Banihani, R. (2020). AnIoT-based framework for early identification and monitoring of COVID-19 cases. Biomedical Signal Processing and Control, 62, art. no. 102149. https://doi.org/10.1016/j.bspc.2020.102149
[20] Ajmera, K. M. (2022). Adverse events profile of COVID-19 preventative strategies. Encyclopedia, 2 (1), pp. 457–465. https://doi.org/10.3390/encyclopedia2010028
[21] Chen, Y.-C., Lu, P.-E., Chang, C.-S., Liu, T.-H. (2020). A time-dependent SIR model for COVID-19 with undetectable infected persons. IEEE Transactions on Network Science and Engineering, 7 (4), art. no. 9200529, pp. 3279–3294. https://doi.org/10.1109/TNSE.2020.3024723
[22] Alazab, M., Awajan, A., Mesleh, A., Abraham, A., Jatana, V., Alhyari, S. (2020). COVID-19 prediction and detection using deep learning. International Journal of Computer Information Systems and Industrial Management Applications, 12, pp. 168–181.
[23] Hilton, J., Keeling, M.J. (2020). Estimation of country-level basic reproductive ratios for novel coronavirus (sars-cov-2/covid-19) using synthetic contact matrices. PLoS Computational Biology, 16 (7), art. no. e1008031. https://doi.org/10.1371/journal.pcbi.1008031
[24] Asteris, P.G., Douvika, M.G., Karamani, C.A., Skentou, A.D., Chlichlia, K., Cavaleri, L., Daras, T., Armaghani, D.J., Zaoutis, T.E. (2020). A novel heuristic algorithm for the modeling and risk assessment of the covid-19 pandemic phenomenon. CMES – Computer Modeling in Engineering and Sciences, 124 (3). https://doi.org/10.32604/cmes.2020.013280
[25] Shorfuzzaman, M., Hossain, M.S. (2020). MetaCOVID: A siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients. Pattern Recognition, 113, art. no. 107700. https://doi.org/10.1016/j.patcog.2020.107700
[26] Vardhan, S., Sahoo, S.K. (2020). In silico ADMET and molecular docking study on searching potential inhibitors from limonoids and triterpenoids for COVID-19. Computers in Biology and Medicine, 124, art. no. 103936. https://doi.org/10.1016/j.compbiomed.2020.103936
[27] Burdick, H., Lam, C., Mataraso, S., Siefkas, A., Braden, G., Dellinger, R.P., McCoy, A., Vincent, J.-L., Green-Saxena, A., Barnes, G., Hoffman, J., Calvert, J., Pellegrini, E., Das, R. (2020). Prediction of respiratory decomposition in Covid-19 patients using machine learning: The READY trial. Computers in Biology and Medicine, 124, art. no. 103949. https://doi.org/10.1016/j.compbiomed.2020.103949
[28] Ray, Samrat. (2021). How can we learn from our mistakes during COVID-19? Circular economy in India through Biogas economics. Economic Environment. https://doi.org/10.36683/2306-1758/2021-2-36/59-65
[29] Ray, Samrat, Dossou, Yédéhou Léandre. (2021). How entrepreneurial university model is changing the Indian COVID–19 fight? Entrepreneur's Guide. 14, pp. 153–162. https://doi.org/10.24182/2073-9885-2021-14-3-153-162
[30] Nguyen, C.T., Saputra, Y.M., Huynh, N.V., Nguyen, N.-T., Khoa, T.V., Tuan, B.M., Nguyen, D.N., Hoang, D.T., Vu, T.X., Dutkiewicz, E., Chatzinotas, S., Ottersten, B. (2020). A comprehensive survey of enabling and emerging technologies for social distancing – Part I: Fundamentals and enabling technologies. IEEE Access, 8, art. no. 9172058, pp. 153479–153507. https://doi.org/10.1109/ACCESS.2020.3018140

[31] Natarajan, A., Su, H.-W., Heneghan, C. (2020). Assessment of physiological signs associated with COVID-19 measured using wearable devices. NPJ Digital Medicine, 3 (1), art. no. 156. https://doi.org/10.1038/s41746-020-00363-7
[32] Wang, Z., Liu, Q., Dou, Q. (2020). Contrastive cross-site learning with redesigned net for COVID-19 CT classification. IEEE Journal of Biomedical and Health Informatics, 24 (10), art. no. 9194240, pp. 2806–2813. https://doi.org/10.1109/JBHI.2020.3023246
[33] Hazarika, B.B., Gupta, D. (2020). Modelling and forecasting of COVID-19 spread using wavelet-coupled random vector functional link networks. Applied Soft Computing Journal, 96, art. no. 106626. https://doi.org/10.1016/j.asoc.2020.106626
[34] Al-Waisy, A.S., Al-Fahdawi, S., Mohammed, M.A., Abdulkareem, K.H., Mostafa, S.A., Maashi, M.S., Arif, M., Garcia-Zapirain, B. (2020). COVID-CheXNet: Hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images. Soft Computing. https://doi.org/10.1007/s00500-020-05424-3
[35] Robson, B. (2020). Bioinformatics studies on a function of the SARS-CoV-2 spike glycoprotein as the binding of host sialic acid glycans. Computers in Biology and Medicine, 122, art. no. 103849. https://doi.org/10.1016/j.compbiomed.2020.103849
[36] Mohammed, M.A., Abdulkareem, K.H., Garcia-Zapirain, B., Mostafa, S.A., Maashi, M.S., Al-Waisy, A.S., Subhi, M.A., Mutlag, A.A., Le, D.-N. (2021). A comprehensive investigation of machine learning feature extraction and classification methods for automated diagnosis of COVID-19 based on X-ray images. Computers, Materials and Continua, 66 (3), pp. 3289–3310. https://doi.org/10.32604/cmc.2021.012874
[37] Santosh, K.C. (2020). COVID-19 prediction models and unexploited data. Journal of Medical Systems, 44 (9), art. no. 170. https://doi.org/10.1007/s10916-020-01645-z
[38] Shorten, C., Khoshgoftaar, T.M., Furht, B. (2021). Deep learning applications for COVID-19. Journal of Big Data, 8 (1), art. no. 18. https://doi.org/10.1186/s40537-020-00392-9
[39] Scopus database: www.scopus.com (Data accessed till 26th September 2021)
[40] VOSviewer download website: https://www.vosviewer.com/download
[41] By: IBM Cloud Education. (n.d.). *What are recurrent neural networks?* IBM. Retrieved December 12, 2021, from https://www.ibm.com/cloud/learn/recurrent-neural-networks#toc-what-are-r-btVB33l5
[42] Prabhakaran, S. (2021, November 22). *Arima model – complete guide to time series forecasting in python: ML+*. Machine Learning Plus. Retrieved December 12, 2021, from https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/
[43] https://www.kaggle.com/hiramcho/xbnet-regressor-for-covid-19-data
[44] Rajeswari, K., Sushma Vispute, Vighnesh Pathrikar, Akshay Siddannavar, Tejas Podutwar, Akash Mandana. (2021). Research on various time forecasting algorithms for predicting Covid-19 cases. International Journal of Engineering Research & Technology (IJERT), ISSN: 2278–0181, 10 (12).
[45] Vinay Kumar Reddy Chimmula, Lei Zhang, (2020). "Time series forecasting of COVID-19 transmission in Canada using LSTM networks", Chaos, Solitons and Fractals 135 (2020) 109864 https://doi.org/10.1016/j.chaos.2020.109864
[46] Amulya Maitre, Sushma Vispute, Rajeswari Kannan, (2021). Time series analysis for understanding the vaccination rate using ARIMA. International Journal of Engineering Research & Technology (IJERT), 10 (7), ISSN. 2278–0181.
[47] https://machinelearningmastery.com/xgboost-for-regression/
[48] https://pythonrepo.com/repo/tusharsarkar3-XBNet
[49] World Health Organization. (n.d.). Coronavirus. World Health Organization. Retrieved August 20, 2022, from https://www.who.int/health-topics/coronavirus#tab=tab_3

[50] Kumar, R.L., Khan, F, Din, S, Band, S.S., Mosavi, A., Ibeke, E. (2021). Recurrent neural network and reinforcement learning model for COVID-19 prediction. Front. Public Health 9:744100. https://doi.org/10.3389/fpubh.2021.744100
[51] Podder, P., Khamparia, A., Mondal, M. R. H., Rahman, M. A., Bharati, S. (2021). Forecasting the spread of COVID-19 and ICU requirements. Int. J. Onl. Eng., 17 (05), pp. 81–99. https://doi.org/10.3991/ijoe.v17i05.20009
[52] Ali, L. R., Jebur, S. A., Jahefer, M. M., Shaker, B. N. (2022). Employing transfer learning for diagnosing COVID-19 disease. Int. J. Onl. Eng., 18 (15), pp. 31–42. https://doi.org/10.3991/ijoe.v18i15.35761
[53] Pathrikar, V., Podutwar, T., Siddannavar, A., Mandana, A., Rajeswari, K., Vispute, S.R. (2022). "Forecasting Diurnal Covid-19 Cases for Top-5 Countries Using Various Time-series Forecasting Algorithms," 2022 International Conference on Emerging Smart Computing and Informatics (ESCI), 2022, pp. 1–6. https://doi.org/10.1109/ESCI53509.2022.9758373

6 Authors

K. Rajeswari, Pimpri Chinchwad College of Engineering, Pune, India.
Sushma Vispute, Pimpri Chinchwad College of Engineering, Pune, India.
Amulya Maitre, Pimpri Chinchwad College of Engineering, Pune, India.
Reena Kharat, Pimpri Chinchwad College of Engineering, Pune, India.
Amruta Aher, Pimpri Chinchwad College of Engineering, Pune, India.
N. Vivekanandan, Pimpri Chinchwad College of Engineering, Pune, India.
Renu Kachhoria, Pimpri Chinchwad College of Engineering, Pune, India.
Swati Jaiswal, Pimpri Chinchwad College of Engineering, Pune, India.

Article submitted 2023-02-23. Resubmitted 2023-04-03. Final acceptance 2023-04-03. Final version published as submitted by the authors.