International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

# Designing a Machine Learning Based Prediction Model for Covid-19 in Ethiopia

Ashenafi Tulu[1], Jemal Abate[1], Vijayshri Khedkar[2], Fikirte Zemene[1]

**Submitted**: 25/09/2023 **Revised**: 16/11/2023 **Accepted**: 27/11/2023

**Abstract:** A global health crisis resulted due to COVID 19, and the medical industry is seeking support of novel technologies to monitor and control disease spread. This study aims to design machine learning-based prediction model for COVID-19 outbreaks in Ethiopia. Large-scale data analytics were used to collect information from the Ethiopian Ministry of Health and WHO reports between March 13, 2020 up to May 16, 2022. The dataset included variables like the number of new cases, tolls and recoveries data from Ethiopia's ten regional states and two administrative cities. Various splitting percentage of the dataset is used as a training set to train the model and a comparison was made for the prediction results of the models. The researchers utilized the AR, ARIMA, and SARIMA models to predict the data accurately, and the best Spearman correlation was used to determine which model was better suited for prediction. It was observed that the ARIMA model performed best at identifying cases that are dead, confirmed, and recovered. For cases that are dead, confirmed, and recovered, respectively, the Spearman correlation ratings for ARIMA were 1.00, 1.00, and 0.93. With less datasets, the AR model did well as well, but the SARIMA model excelled with larger datasets. The proposed machine learning-based prediction model is a useful tool for controlling and monitoring the spread of COVID-19 in Ethiopia.

***Keywords:*** *Machine Learning, time series approach, Prediction, COVID-19, Ethiopia, Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA) model, Seasonal Autoregressive Integrated Moving Average (SARIMA).*

## 1. Introduction

According to WHO (World health organization) in 2019, the disease COVID-19 is caused by SARS-CoV-2 which is a type of corona virus. Wuhan, People's Republic of China reported WHO first of this novel virus on 31 December 2019, following a report of a huge number of cases of viral pneumonia. By January 2021, more than ninety five million cases were documented across the globe, with a mortality rate of 2% of all reported cases (Clemente-Suárez et al., 2021) [6]. Its fast pandemic expansion was a global worry and an acute threat to the global economy besides public health. To prevent the sickness from spreading countries started restricting citizens in home and socialization stopped to help stop the spread of virus (Kolozsvári et al., 2021) [9],[15].

Ethiopia faces a more difficult and perilous task than many other countries, because disrupting daily economic life would jeopardize the livelihoods of tens of millions of people already living in vulnerable situations. COVID-19 infections are so widespread in Ethiopia, the doctor pointed out that confirming the impact of COVID-19 will be challenging (Piccialli et al., 2021) [13]. Ethiopia's surveillance system is insensitive to changes in disease trends, and that a lack of testing techniques and facilities may have impeded the country's ability to discover infections quickly and accurately track index cases (Nigussie, 2021) [11].

To stop the virus from spreading and to help the economy, officials rapidly put in place restrictive measures such as closing schools, universities, and religious organizations, as well as prohibiting sporting events and mass gatherings. They've also increased screening and bolstered the health sector's ability to respond to the issue (Gaye et al., 2020) [8]. According to Ethiopia's health minister, in the battle against the pandemic, the ability to gauge the disease's rate of spread is essential. Aiding governments in public health policy development to handle the pandemic's implications is possible by knowing the limit of the pandemic's spread at any particular time [2].

Various studies are suggesting to focus on developing new methodologies to assess the likelihood of an outbreak, which is crucial for developing efficient disease prevention plans.

AI approaches have been found to play a vital role in reducing the effect of the virus spread [1],[3], attempted to design and construct Exponential Smoothing Model, and stated Double Exponential Smoothing method was appropriate in forecasting the future number of COVID-19 cases in Ethiopia. Perone (2022) [12], predict, as of 20 August 2021, the short- to mid-term cumulative mortality from COVID-19 in 12 severely affected large countries worldwide. The COVID-19 dataset from the World in Data

[1] *College of Computing and Informatics, P.O. Box 138, Haramaya University, Dire Dawa, Ethiopia*
*ORCID ID: 0000-0002-0349-7505*
*ORCID ID: 0000-0002-7903-6940*
[2] *Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India*
*ORCID ID: 0000-0001-6704-4823*
** Corresponding Author Email: vijayshri.khedkar@sitpune.edu.in*

was utilized to extract the data for the analysis. A seasonal and non-seasonal autoregressive integrated moving average (SARIMA) estimation was performed. Figa (2020) [7], has applied the Autoregressive Integrated Moving Average (ARIMA) modeling approach for projecting COVID-19 prevalence in selected East African countries. Argawu (2020) [4], developed a OLS model to predict Number of new laboratory tests and number of new cases from Addis Ababa city.

In Ethiopia, there is still a need for technology and analytical methods that are capable of predicting the likelihood of an outbreak, which is crucial for developing efficient disease management plans. AI is likely to play a key role in the worldwide is capable of predicting the likelihood of an outbreak, which is crucial for developing efficient disease management plans. To the best of the researchers' knowledge, no research has been done considering the cases of all regions in Ethiopia. This study will contribute by designing and developing an AI-based prediction model for combating COVID-19 using the data collected from ten regional states and two administrative cities in Ethiopia.

Therefore, an AI-based temporal sequence method was suggested in order to predict the proportion of new instances, death cases and recovered cases with respect to the total population for the ten regional and two administrative cities of Ethiopia. Various time series approaches, namely Autoregressive (AR), ARIMA and Seasonal Autoregressive Integrated Moving Average (SARIMA) were used to develop the system. When it comes to evaluation metrics, the Spearman correlation was put to use to measure the efficacy of each time series model. The evaluation result shows that when the size of the dataset increases the performance of the ARIMA is higher than the others and its performance will be low if it is performed in a few datasets. Additionally, the SARIMA model can perform better with few dataset amounts.

## 2. Materials and Methods

### 2.1. Research Design

In this project design science research is applied, because it is a method for defining, explaining, and forecasting the consequences of technology on individuals, organizations, or society. Concern is the methodical accumulation of information about and with design as a deliberate, intellectual, and creative problem-solving activity (Carstensen & Bernhard, 2019) [5]. There are the steps and processes to follow in the design science research methods. These processes include: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication (vom Brocke et al., 2020) [14]. Design science research involves method of data collection and data set preparation, design and development of AI analytics system to fight against covid19.

### 2.2. Data source and Method of Data Collection

As a result of the outbreak, Ethiopia Ministry of Health and WHO have created publicly available datasets regarding COVID-19 data for analysis, research and for awareness creation. The data were collected from Ethiopia Ministry of Health and WHO for the purpose that support the model prediction function. The dataset used were a total of 2100 cases which contains number of date of recorded, new cases, new deaths, and the number of new recoveries from ten regional state and two administrative cities of Ethiopia. Out of 2100 dataset the daily recovery column contains 510 instances, whereas the daily confirmed cases and daily death cases accounts 795 instances each. To trains the system for cases estimator the researcher encoded the data in the form of a data frame, organized it by the date, and then create separate frames for the confirmed, recovered cases and death. The dataset file contains the information of daily confirmed, recovered cases and death.

### 2.3. Tools Used

Various software, hardware, open-source libraries and programming languages are used as development tools in this study. Fbprophet library is an open-source library which is created by Facebook and used for forecasting. TensorFlow software libraries were used to develop the system. It's used because TensorFlow eases the process of acquiring data, training models, serving predictions, and refining future results. Among the programming language tools python is used to develop the prototype. Another Python module called pandas provides a dedicated class for time series objects.

### 2.4. Developed Architecture

The architecture of the system is designed to perform various activities such as to forecast the number of COVID-19 positive cases, show a survival rate based on distribution, select the best epidemic model based on observable new cases after training on the history of daily new confirmed cases. The architecture of this study will have the components namely: data collection, data preprocessing, feature engineering, model selection, model training, model tuning and model implementation.

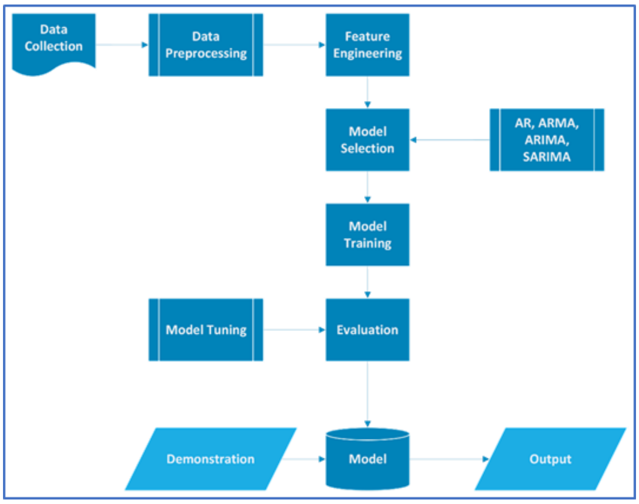


**Fig. 1** System Architecture

### 2.4.1. Data collection:

Collection of data from reliable sources such as the Ethiopian Ministry of Health and WHO reports on COVID-19 cases, deaths, recoveries, and other relevant variables is done. The dataset used for confirmed and deaths ranges from March 13, 2020 up to May 16, 2022. Whereas, the data used for recovered cases ranges from March 13, 2020 up to August 04, 2021.

```
df = pd.read_csv('last_data/data_846.csv',parse_dates=['Date_of_Observation'])
df.head()
```

| | Date_of_Observation | Number_of_new_Cases | Number_of_new_Recovered | Number_of_new_Deaths |
|---|---|---|---|---|
| 0 | 2020-03-13 | 1 | 0.0 | 0 |
| 1 | 2020-03-14 | 0 | 0.0 | 0 |
| 2 | 2020-03-15 | 3 | 0.0 | 0 |
| 3 | 2020-03-16 | 1 | 0.0 | 0 |
| 4 | 2020-03-17 | 1 | 0.0 | 0 |

**Fig. 1** The First 5 Sample Dataset

### 2.4.2. Data pre-processing

The data is preprocessed by eliminating missing values and outliers, and converting it to a suitable format for analysis. Furthermore, relevant characteristics such as daily new cases, deaths, and recoveries are extracted at this stage. After loading the data from the CSV file, the describe() method is applied to the data frame to obtain a basic statistical summary of the data. Next, the data frame's columns are renamed using a dictionary object, with the columns 'Date_of_Observation', 'Number_of_new_Cases', 'Number_of_new_Recovered', and 'Number_of_new_Deaths' being renamed to 'Date', 'Confirmed', 'Recovered', and 'Deaths', respectively. The screenshot below shows the output of the summary result.

```
df = pd.read_csv('last_data/data_846.csv',parse_dates=['Date_of_Observation'])
df.describe()
```

| | Number_of_new_Cases | Number_of_new_Recovered | Number_of_new_Deaths |
|---|---|---|---|
| count | 795.000000 | 510.000000 | 795.000000 |
| mean | 592.441509 | 517.662745 | 9.449057 |
| std | 723.473203 | 571.944325 | 9.907550 |
| min | 0.000000 | 0.000000 | 0.000000 |
| 25% | 99.000000 | 68.500000 | 2.000000 |
| 50% | 404.000000 | 340.000000 | 7.000000 |
| 75% | 785.000000 | 801.250000 | 13.000000 |
| max | 5185.000000 | 4493.000000 | 49.000000 |

**Fig. 3** Summary of statistics

### 2.4.3. Feature engineering

Feature engineering is the process of adding new variables and features to the final dataset used to train a model utilizing historical row data (Lazzeri, 2020) [10]. Time series data can be divided into two types: stationary and non-stationary. Stationarity is an important property because some models work well with stationary data. Time series data, on the other hand, is frequently non-stationary. As a result, we must understand how to identify non-stationary time series and transform them using various techniques, such as differencing. (Yugesh Verma, 2022) [16]. The two-way correlation between the current observation and previous observations has been summarized using the ACF (Autocorrelation Function). The line of code below creates ACF graphs for 40 delays.

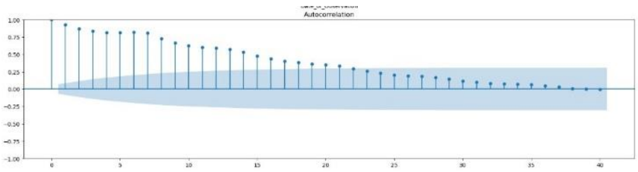


**Fig. 2** Two-way correlation for confirmed cases

Non-stationary data appear to be highly correlated with earlier observations, suggesting that time component still plays a significant role, ACF for non-stationary data decreases to 0 quite slowly. The diagram above shows the ACF of the original time series data, which decreases slowly thus very likely to be non-stationary.

### 2.4.4. Model selection:

Select appropriate machine learning models that can handle time series data, such as AR, ARIMA, and SARIMA.

**A)** **AR Model:** In an auto regression model, we use a linear combination of the variable's historical values to forecast the variable of cases. The word auto regression denotes that the variable is being regressed against itself. Thus, an autoregressive model of order can be written as follows in equation 1:

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon_{t,} \quad \ldots \text{equation 1}$$

Where white noise is denoted by $\varepsilon_t$ . A multiple regression with lagged values of $y_t$ as predictors can be analogous. Model refers to it as an $AR(p)$ model, an autoregressive model having order $p$.

**B) ARIMA Model:** An ARIMA (p, q) model is the combination of AR(p) and MA(q) models that can be used to model univariate time series. The formula for an ARIMA (p, d, q) model is written in following equation 2:

$$y'_t = c + \emptyset_1 y'_{t-1} + \cdots + \emptyset_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_{t,} \ \ldots \text{ equation 2}$$

Where: the differenced series is $y'_t$. The right-hand side has “predictors” including both lagged values of $y_t$ and lagged errors. This is an ARIMA (p, d, q) model, where p denotes order of the autoregressive part; d denotes degree of first differencing involved; q denotes order of the moving average part.

**C) SARIMA Model:** In case of non-stationary data and non-seasonal data, the ARIMA model is used. The Seasonal ARIMA (SARIMA) model will depict the Seasonal differencing of suitable order which is employed in this model to avoid non-stationarity from the series.

### 2.4.5. Model training and Model tuning

To train the selected models, preprocessed data is used to train the selected models, which are then evaluated for performance using metrics like Spearman correlation. The models are fine-tuned based on the evaluation results to improve their accuracy and performance. Various packages are utilized to forecast the number of confirmed, death, and recovered cases for the coronavirus outbreak. Each case category's models are initialized with a defined interval width and then fitted to the given time series data. Future data frames are generated for each category, which are then used to predict the number of cases using the fitted models.

### 2.4.6. Model Implementation

The final AI-based prediction model for COVID-19 in Ethiopia has been implemented, which can aid health officials and policymakers in making informed decisions and taking preventative measures to combat the pandemic. The implementation utilizes time series forecasting, which involves examining previous observations of a random variable to create a model that accurately represents the underlying relationship and its patterns. Time series forecasting is particularly useful when there is limited information about the underlying data-generating distribution or when no explanatory model can effectively relate the prediction variable to other explanatory factors. Over the past few decades, a considerable amount of research has been devoted to developing and improving time series forecasting models. This section provides a brief overview of the forecasting approaches used such as: AR, ARIMA, and SARIMA.

### A. Forecasting using the AR

The AR model was utilized for forecasting the outcomes of confirmed cases, death cases, and recovered cases. To compare the predicted values with the known values, a plot was created that displays the values of the confirmed cases, death cases, and recovered cases. The following plot on Figure 3, provides a visual comparison between the current and predicted data using AR.

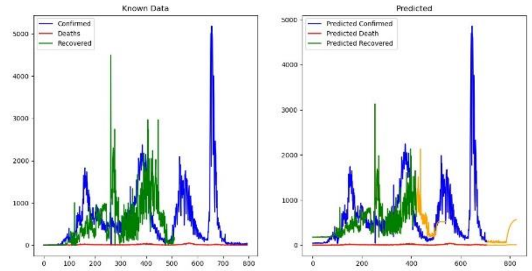


**Fig. 5** Comparison between the current and predicted data (AR).

Fig. 4 is composed of two subplots, with the left side showing the known values of the data, and the right side reflecting the forecasted values. This visual representation allows for a clear comparison between the actual and predicted values for the different cases. Upon examining Figure 5, it can be observed that the predicted trend is more aligned with the known or test data. This suggests that the AR model's forecasts for new cases, death cases, and recovered cases are relatively accurate and closely follow the observed patterns in the data.

### B. Forecasting the Cases using the ARIMA Model

below Fig. shows the observed cases versus prediction of the coronavirus new cases, new death and recovered cases of ARIMA.

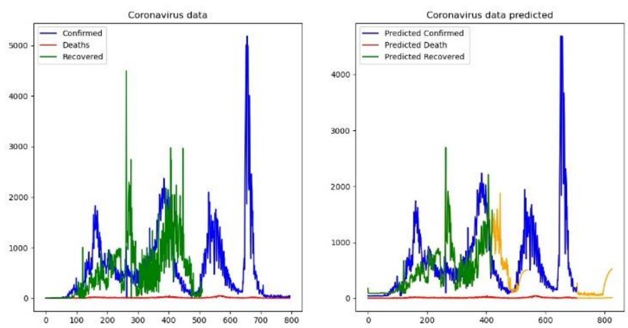


**Fig.6** Comparison between the current and predicted data (ARIMA)

The predictions generated by the ARIMA model are visually appear to be accurate, as the plotted data points closely align with the observed cases. Furthermore, the width of the Result of Spearman correlation interval has decreased, indicating that the model has become more confident in its predictions. The results from the ARIMA model, as shown in Fig.6, suggest that it is capable of producing reliable predictions for new cases, new deaths, and recovered cases of coronavirus.

### C. Forecasting the Cases using SARIMA Model

The predictions using the SARIMA model is also plotted to show the prediction and the known data in order to compare. Fig. 7, indicate the plot of the data for the confirmed cases and the forecasted cases.

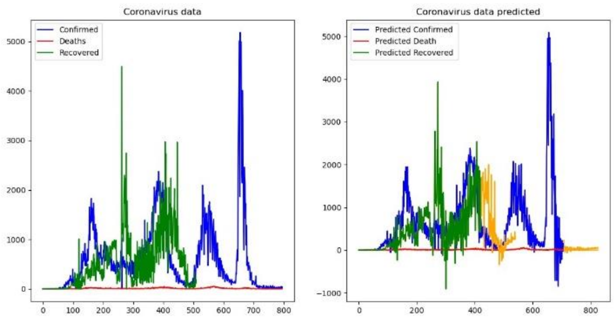


**Fig. 7** Comparison between the current and predicted data (SARIMA)

Upon examining the plot in figure Fig. 7, it is evident that the dataset used in this study contains a high number of confirmed cases compared to recovered cases. As a result, the SARIMA model has achieved high accuracy in predicting the confirmed cases. The plotted data points for the predicted cases closely align with the observed cases, indicating that the SARIMA model is capable of generating reliable predictions. The results from the SARIMA model, as depicted in Fig. , suggest that it is effective in forecasting the confirmed cases of coronavirus. The close alignment between the predicted and observed data points indicates that the SARIMA model is capturing the underlying patterns and trends in the data, including the seasonal fluctuations.

## 3. Evaluation

To evaluate the model's performance, the Spearman correlation evaluation metrics was used. Various splitting percentage of the dataset is used as a training set to train the model and a comparison was made for the prediction results of the models. When one of the variables is a perfect monotone function of the other, a perfect Spearman correlation of +1 or -1 happens when there are no repeated data values. When observations have a similar and different rank between the two variables, the Spearman correlation is strong (correlation of 1), and low (Correlation of -1) respectively.

**Table 1** Evaluation Result of Spearman correlation

| **S. No.** | **Methods** | **Confirmed score** | | | **Death score** | | | **Recovered score** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ***100%*** | ***80%*** | ***60%*** | ***100%*** | ***80%*** | ***60%*** | ***100%*** | ***80%*** | ***60%*** |
| 1. | AR model | 0.88 | 0.92 | 0.93 | 0.78 | 0.89 | 0.89 | 0.76 | 0.89 | 0.87 |
| 2. | ARIMA model | 1.00 | 0.99 | 0.98 | 1.00 | 0.99 | 0.99 | 0.93 | 0.95 | 0.94 |
| 3. | SARIMA model | 0.95 | 0.96 | 0.96 | 0.88 | 0.90 | 0.89 | 0.81 | 0.84 | 0.85 |

Each method has scored different score for confirmed, death and recovered cases. As a result, the ARIMA model has the highest score for predicting the confirmed, deaths and the recovered cases. The dataset used in this study has high number of confirmed and the death cases as compared to recovered cases. The evaluation result shows that when the size of the dataset increases the performance of the ARIMA is higher than the others and its performance will be low if it is performed in a few datasets. Additionally, the AR model can perform better with few dataset amounts.

## 4. Result and Discussion

The COVID-19 pandemic has caused an unprecedented global health crisis, prompting the need for effective measures to combat its spread. This study aims to develop a prediction model using artificial intelligence techniques to aid in the fight against COVID-19. Specifically, the study evaluates the performance of three-time series forecasting models (AR, ARIMA, and SARIMA) in predicting the number of confirmed cases, deaths, and recoveries.

To evaluate the performance of these models, the study employs the Spearman correlation coefficient as an evaluation metric, which measures the strength and direction of the monotonic relationship between the predicted and actual values. The results of the study are presented in Table 1 and show that all three models perform well in predicting COVID-19 metrics, with correlation coefficient values ranging from 0.76 to 1.00.

According to the results, the ARIMA model performs the best among all three models, with perfect positive correlation coefficients of 1.00 for both confirmed and death cases. The SARIMA model also performs well in predicting recovered cases, with a correlation coefficient of 0.81. On the other hand, the AR model has the lowest correlation coefficients among all three models, indicating that it may not be as effective as the other two models in capturing the underlying patterns and trends in the data.

The effect of data splitting percentage on the performance of the models was also observed. The results show that as the size of the dataset increases, the performance of the ARIMA model tends to be higher than the other models. Specifically, when trained with a larger percentage of the dataset (100% or 80% data), the ARIMA model achieves perfect positive correlation coefficients of 1.00 for confirmed and death cases, indicating its high accuracy in predicting the number of cases and deaths related to COVID-19. However, when trained with a smaller percentage of the dataset (60% data), the correlation coefficients of the ARIMA model decrease slightly to 0.93 for confirmed cases and 0.94 for death cases, suggesting a relatively lower performance compared to the other two models.

The high correlation coefficients obtained by the ARIMA and SARIMA models suggest that these models are effective in capturing the trends and patterns of the COVID-19 outbreak. Specifically, the perfect correlation coefficient values obtained by the ARIMA model for confirmed and death cases indicate that it can accurately predict the number of cases and deaths related to the COVID-19 outbreak. This could be useful in helping governments and healthcare organizations to better prepare for the impact of the outbreak and take appropriate measures to mitigate its effects.

The results of this study suggest that artificial intelligence techniques, particularly the ARIMA model, can be useful in predicting the impact of the COVID-19 outbreak. By providing accurate predictions of confirmed cases, deaths, and recoveries, governments and healthcare organizations can take proactive measures to manage the pandemic effectively. Furthermore, these results demonstrate the potential of artificial intelligence techniques in supporting public health policy, particularly during times of crisis.

## 5. Conclusion

In this study, an AI-based model was developed to predict COVID-19 outbreak analytics using AR, ARIMA, and SARIMA models. The models were evaluated using Spearman correlation evaluation metrics, and the results showed that the ARIMA model had the highest performance in predicting confirmed, death, and recovered cases. The Spearman correlation scores for ARIMA were 1.00, 1.00, and 0.93 for confirmed, death, and recovered cases, respectively. The AR model also showed good performance with fewer datasets, while the SARIMA model performed well when the dataset size was larger.

The results of this study have important implications for public health policy, demonstrating the potential of artificial intelligence techniques in supporting effective management of pandemics. However, it is important to note that further research is needed to validate these findings and to identify any potential limitations of the models. Nevertheless, this study provides a strong foundation for future research in this area and highlights the importance of using advanced technologies to combat global health crises.

Furthermore, the study provides important insights into the effectiveness of different time series forecasting models so as to predict the COVID-19 outbreak. The high performance of ARIMA and SARIMA models suggests their potential as useful tools in predicting the impact of the outbreak and informing decision-making. However, it is crucial to conduct further research to explore the influence of other factors on the outbreak. Additionally, the study highlights the importance of considering the dataset size and splitting percentage when utilizing these models, as it can impact their performance. Overall, this study contributes to the growing body of literature on COVID-19 prediction modeling and underscores the significance of leveraging advanced technologies for addressing global health challenges. So, it is essential to continue exploring and refining predictive models to support effective decision-making during pandemics.

## 6. Recommendation

The study highlights the potential of machine learning-based time series methods for predicting COVID-19 data in Ethiopia. Therefore, it is recommended that the Ethiopian government and other key stakeholders continue to leverage artificial intelligence and machine learning technologies for combatting the COVID-19 pandemic in the country.

The study recommends the use of the ARIMA model for predicting daily confirmed cases and daily recovered patients, and for predicting daily death cases. However, the study also highlights the importance of choosing the right time series model for different variables and dataset sizes. Therefore, it is recommended that further research be conducted to explore other machine learning methods that may perform better for different variables.

The study's findings can be useful in guiding public health policies and interventions for combating the COVID-19 pandemic in Ethiopia and other similar settings. Therefore, it is recommended that the Ethiopian government and other key stakeholders use the results of this study to inform and guide their decision-making processes related to COVID-19 control measures.

## 7. Ethical Consideration

This study will not use personal information such as name, gender, identification number and other information that may describe person's identity. Any COVID-19 patients' information which will harm the individual under investigation will not be disclosed without the consent of the individuals. All reference materials and any other help full information which will be used for this project will be cited and acknowledged.

**References**


[1] Abebe, T. (2021). Forecasting the Number of Coronavirus (COVID-19) Cases in Ethiopia Using Exponential Smoothing Times Series Model. International Journal of Biomedical Engineering and Clinical Science, 10(11648), 1–6. https://doi.org/10.11648/j.ijbecs.20210701.11

[2] Ahmed, M. A., Ly, B. A., Millimouno, T. M., Alami, H., Faye, C. L., Boukary, S., Accoe, K., Van Damme, W., Put, W. Van De, Criel, B., & Doumbia, S. (2020). Willingness to comply with physical distancing measures against COVID-19 in four African countries. BMJ Global Health, 5(9), 1–4. https://doi.org/10.1136/bmjgh-2020-003632

[3] Al-qaness, M. A. A., Saba, A. I., Elsheikh, A. H., Elaziz, M. A., Ibrahim, R. A., Lu, S., Hemedan, A. A., Shanmugan, S., & Ewees, A. A. (2021). Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil. Process Safety and Environmental Protection, 149, 399–409. https://doi.org/https://doi.org/10.1016/j.psep.2020.11.007

[4] Argawu, A. S. (2020). Regression Models for Predictions of COVID-19 New Cases and New Deaths Based on May/June Data in Ethiopia. Endocrine, 9(May), 6.

[5] Carstensen, A. K., & Bernhard, J. (2019). Design science research–a powerful tool for improving methods in engineering education research. European Journal of Engineering Education, 44(1–2), 85–102. https://doi.org/10.1080/03043797.2018.1498459

[6] Clemente-Suárez, V. J., Navarro-Jiménez, E., Moreno-Luna, L., Saavedra-Serrano, M. C., Jimenez, M., Simón, J. A., & Tornero-Aguilera, J. F. (2021). The impact of the covid-19 pandemic on social, health, and economy. Sustainability (Switzerland), 13(11), 1–25. https://doi.org/10.3390/su13116314

[7] Figa, R. (2020). Stochastic modelling for predicting COVID-19 prevalence in East Africa Countries. Infectious Disease Modelling, 5.

[8] Gaye, B., Khoury, S., Cene, C. W., Kingue, S., N'Guetta, R., Lassale, C., Baldé, D., Diop, I. B., Dowd, J. B., Mills, M. C., & Jouven, X. (2020). Socio-demographic and epidemiological consideration of Africa's COVID-19 response: what is the possible pandemic course? Nature Medicine, 26(7), 996–999. https://doi.org/10.1038/s41591-020-0960-y

[9] Kolozsvári, L. R., Bérczes, T., Hajdu, A., Gesztelyi, R., Tiba, A., Varga, I., Al-Tammemi, A. B., Szőllősi, G. J., Harsányi, S., Garbóczy, S., & Zsuga, J. (2021). Predicting the epidemic curve of the coronavirus (SARS-CoV-2) disease (COVID-19) using artificial intelligence: An application on the first and second waves. Informatics in Medicine Unlocked, 25(July). https://doi.org/10.1016/j.imu.2021.100691

[10] Lazzeri, F. (2020). Machine learning for time series forecasting with python. Machine Learning for Time Series Forecasting with Python, 1–206. https://doi.org/10.1002/9781119682394

[11] Nigussie, H. (2021). The Coronavirus Intervention in Ethiopia and the Challenges for Implementation. Frontiers in Communication, 6(May), 1–12. https://doi.org/10.3389/fcomm.2021.562512

[12] Perone, G. (2022). Using the SARIMA Model to Forecast the Fourth Global Wave of Cumulative Deaths from COVID-19: Evidence from 12 Hard-Hit Big Countries. Econometrics, 10(2). https://doi.org/10.3390/econometrics10020018

[13] Piccialli, F., di Cola, V. S., Giampaolo, F., & Cuomo, S. (2021). The Role of Artificial Intelligence in Fighting the COVID-19 Pandemic. Information Systems Frontiers, 1467–1497. https://doi.org/10.1007/s10796-021-10131-x

[14] vom Brocke, J., Hevner, A., & Maedche, A. (2020). Introduction to Design Science Research. Springer International Publishing. https://doi.org/10.1007/978-3-030-46781-4_1

[15] WHO. (2019). Coronavirus disease (COVID-19). https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19

[16] Yugesh Verma. (2022, May 24). Quick way to find p, d and q values for ARIMA. https://analyticsindiamag.com/quick-way-to-find-p-d-and-q-values-for-arima/

[17] Sharma, R., Dhabliya, D. Attacks on transport layer and multi-layer attacks on manet (2019) International Journal of Control and Automation, 12 (6 Special Issue), pp. 5-11.