Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Machine Learning-based Decision Support System for Healthcare in the Context of COVID-19: Case Study of Saudi Arabia

Alshammari Budor[1], Kiran Sultan[2], Bassam Zafar[3]

[1]*Department of Computing and Information Systems, King Abdulaziz University, Jeddah, Saudi Arabia.*
[2]*Department of Computer and Information Technology, The Applied College, King Abdulaziz University, Jeddah, Saudi Arabia.*
[3]*Department of Computing and Information Systems, King Abdulaziz University, Jeddah, Saudi Arabia.*
Email: bsalshammari@stu.kau.edu.sa

The COVID-19 pandemic has highlighted the significance of effective healthcare decision-making and resource allocation, posing substantial problems for healthcare systems worldwide, including Saudi Arabia. This work describes the creation of a Machine Learning-based Decision Support System (DSS) designed to improve the management of COVID-19 outcomes, with a focus on hospitalization, recovery, and death rates. The study uses a variety of machine learning methods, including Decision Trees, Linear Regression, Random Forest, and SARIMAX, to examine large datasets of daily confirmed cases, recoveries and fatalities. The Random Forest model outperformed the Linear Regression and SARIMAX models in terms of predicted accuracy. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) figures demonstrate the Random Forest model's superiority, which is particularly good in reflecting the complexity of COVID-19 spread. Furthermore, the study underlines the value of extensive datasets, good feature selection, and model validation in improving prediction accuracy. These findings have significant implications for healthcare practitioners and policymakers, allowing for more informed decision-making and effective resource management during the current pandemic. As a result, the study calls for continuous model refining, multidisciplinary collaboration, and real-time data integration to increase the impact of machine learning applications in healthcare, ultimately leading to better patient outcomes and responsiveness during public health crises.

Keywords: P Machine Learning Decision Support, COVID-19 Resource Management, Prediction Accuracy in Healthcare.

1. Introduction

The dynamic of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infection raised concerns about resource availability across medical systems, including intensive care unit (ICU) healthcare providers, personal protective equipment, total hospital and ICU beds, and mechanical ventilators (Lupei et al., 2022). The COVID-19 pandemic had an immediate and significant social, economic, and health impact, necessitating rapid transformation of quality-focused health systems (Silva et al., 2022).

The COVID-19 pandemic has disrupted daily services because to community-wide mitigating measures implemented by numerous countries (Khan et al., 2021). Due to the minimal possibility of acquiring a vaccine in the near future, global efforts have greatly focused on social distance and complete city and state lockdowns in many instances as the sole solutions to manage the epidemic (Khan et al., 2021). During the first phase of the COVID-19 epidemic, the disease's diagnosis was hampered by the variability in symptoms and imaging findings (with some cases having no signs of fever or radiologic abnormalities) and the severity of disease at the time of presentation (Alsayer et al., 2021). Stratifying illness severity is an important element of patient care; however, during a pandemic, its importance grows and expands to improve patient safety while also maximizing hospital resource use (Wallace et al., 2020).

COVID-19 infection causes systemic and respiratory symptoms in the patient. Systemic symptoms include fever, cough, weariness, sputum production, headache, hemoptysis, acute heart damage, hypoxemia, dyspnea, lymphopenia, and diarrhea (Rothan & Byrareddy, 2020). Respiratory disorders include rhinorrhea, sneezing, sore throats, pneumonia, ground-grass opacities, RNAemia, and acute respiratory distress syndrome (Abbaspour Onari et al., 2021).

Approximately 80% of COVID-19 patients experience moderate symptoms that resolve within two weeks (Gomes, 2020). However, approximately 20% of patients may require hospitalization and other medical treatment (Gomes, 2020). The death rate among severe patients is approximately 13.4% (Wu et al., 2020). Assessing patient risk quantitatively and objectively is crucial for managing patients and allocating medical resources (Gomes, 2020). Non-severe patients can get general quarantine and symptomatic therapy at home or mobile hospitals. Severe patients require immediate transfer to the intensive care unit (ICU) (Wu et al., 2020). Previous studies have summarized the clinical and radiological characteristics of severe COVID-19 patients, but the prognostic usefulness of individual factors remains unclear (Wu et al., 2020).

Various machine learning methods have been presented to forecast the likelihood of acquiring serious complications and fatality (Karthikeyan et al., 2021). This is significant since the number of COVID-19 patients is growing faster than the available services (Karthikeyan et al., 2021). Resource allocation and distribution among patients based on their prognosis is a significant consideration (Karthikeyan et al., 2021).

A Decision Support System (DSS) is a computer-based system that uses data and decision logic to support a human decision-maker. A decision support system does not make any decisions. Instead, technology helps the human decision-maker by analyzing data and providing processed information in an understandable style (Aggarwal et al., 2021). Machine learning (ML) and artificial intelligence (AI) approaches can be used to comprehend patient subgroups, guide clinical decision-making, and improve both operational and patient-centered results (Debnath et al., 2020). Clinical Decision Support Systems (CDSSs) using AI-based models, image processing, and big data analytics can help healthcare professionals with disease prediction, anomaly analysis, patient history modeling, and other clinical tasks (Gangavarapu et al., 2020). A CDSS for predicting COVID-19 infection with diagnostic scans can benefit both healthcare professionals and patients (Mayya et al., 2021). A CDSS for predicting COVID-19 infection with diagnostic scans can benefit both healthcare professionals and patients (Forthmann & Pfleiderer, 2019). Cameras for patient monitoring can enable a contactless X-ray scan procedure (Forthmann & Pfleiderer, 2019).

Researchers face a problem in creating an efficient, quick, and intelligent diagnosing system. Radiologists are employing manual lung infection quantification to detect COVID-19 intensity, whereas AI-based pneumonia is being used (Siddiqui et al., 2021). These algorithms are more efficient at evaluating CT scan results in a shorter time than other methods for identifying early to critical stages of confirmed cases, manual estimation of septic regions on CT scans and X-rays is time-consuming and difficult to evaluate, however, growing lung infections require many CT scan pictures (Siddiqui et al., 2021). AI has been used in health care to anticipate illness spread and construct diagnostic and prognostic models (Syeda et al., 2021). Ye et al. (2020) examined big data, cloud computing, mobile health, and AI as pandemic-fighting technologies, these technologies and data from social media, radiological scans, omics, pharmacological databases, and public health organizations have been utilized to forecast disease (Ye et al., 2020).

Due to their precision, machine learning is employed extensively for predictions. The new online poverty database is a challenge for machine learning (ML) algorithms (Kwekha-Rashid et al., 2023). For instance, choosing the right parameters for model training or the optimum machine learning model for prediction is difficult, researchers used the best machine learning model for the dataset to make predictions (Shinde et al., 2020). Many researchers are using mathematical and machine-learning-based prediction models to predict future epidemic patterns to help governments, health service providers, and society cope with the unprecedented rise in coronavirus infections worldwide (Elarabi et al., 2021). Different machine learning methods assist researchers analyze the trend curve. These may improve the epidemic fight and minimize or eliminate preventive measures, allowing individuals to return to their daily life (Elarabi et al., 2021).

Saudi Arabia, a vast country in terms of land size, is organized into five planning regions, 13 administrative areas, and 118 governorates. It borders five Arabian Gulf countries and a few more Arab countries, with a combined native and foreign population of 35.3 million spread across 2.2 million square kilometres (Salam et al., 2022). This primarily urban country constructed residential, economic, educational, medical, and other infrastructure to foster community living, which accelerates the probability of infection (Salam et al., 2022). The Saudi Vision 2030 proposes substantial structural reforms in the healthcare industry to fulfill the Kingdom's growing demand for healthcare services (Algaissi et al., 2020). Saudi Arabia is one of the countries that could have a significant impact during a global pandemic, as millions of Muslims from all over the world go to Saudi Arabia for Umrah and Hajj, two of the world's greatest mass gatherings (Algaissi et al., 2020).

A Saudi study Elarabi et al. (2021) is based on an analysis of COVID-19 data from Saudi Arabia. It depicts the predicted number of new confirmed cases and deaths from COVID-19 in Saudi Arabia for the following ten days beginning July 8, 2021, during the Hajj season. Machine learning models utilized include Support Vector Machine (SVM), Bayesian Edge (BR), Linear Regression (LR), and Moving Average (MA). Each model makes two predictions: the number of newly infected patients and deaths in the next ten days. The results show that SVM, Bayesian Edge (BR), Linear Regression (LR), and Moving Average (MA) forecasts are all highly accurate, with LR performing particularly well. When combined with the available dataset, BR performs badly in prediction scenarios for new confirmed cases. All models accurately predicted deaths, with SVM coming out on top, followed by Bayesian Edge. It also predicts a rise in confirmed cases under the SVM model scenario to 511,257 on July 17 from 496,516 on July 7 in the actual daily cumulative cases. The death toll will rise to 8,113 on July 17 from 7,921 on July 7, according to actual daily cumulative figures (Elarabi et al., 2021).

In fact, four Saudi systems have recently been implemented to help the diagnosis and management of COVID-19: To begin, the Health Electronics Surveillance Network (HESN) is a web-based e-health solution that is an integrated and flexible system capable of accommodating all Saudi public health programs, and the Public Health Information System (PHIS) is designed to better support public health professionals across the Kingdom (Bangar et al., 2020). During the COVID-19 epidemic, few papers employed machine learning to predict mortality and identify risk variables (Elhazmi et al., 2022). Classical statistical analysis methods used to detect such risk variables are restricted in their capacity to emphasize the effect on outcome caused by potential interactions among these factors (Elhazmi et al., 2022).

Numerous datasets have been created to better understand the COVID-19 epidemic. Predicting COVID-19 infection is challenging due to its high volume, risk variables, and complexity. Factors like demographics, medical history, symptoms, and real-time data updates all contribute to COVID-19 complications (Wynants et al., 2020) . This level of complexity requires advanced computational methods and extensive processing time. Innovative techniques, such as machine learning (ML), are needed to predict the severity of asymptomatic carriers and the death rate associated with documented illnesses (Ghandorh et al., 2024).

The major goal of this work was to create a Machine Learning-based Decision Support System (DSS) suited for Saudi healthcare, with a specific focus on predicting COVID-19-related outcomes such as hospitalization, recovery, and mortality. The study aimed to develop a Machine Learning-based Decision Support System (DSS) specifically designed for healthcare in Saudi Arabia to predict COVID-19-related outcomes, including hospitalization, recovery, and mortality. The study strived to enhance healthcare decision-making processes by providing accurate, timely forecasts of COVID-19 cases, recoveries, and fatalities, thereby supporting effective resource allocation and health management strategies in the context of the ongoing pandemic.

Contributions of the Study

This study greatly contributes to the field of healthcare decision support systems (DSS) by incorporating machine learning approaches to improve COVID-19 management in Saudi Arabia. The study's primary contributions are:

✓ Development of a Machine Learning-based Decision Support System: The study used a customized DSS that employs different machine learning models (Decision Tree, Linear Regression, and Random Forest) to properly forecast COVID-19 outcomes. This helps with real-time decision-making for resource allocation and patient management during the pandemic.

✓ Focus on Comprehensive Datasets: By using a diverse mix of epidemiological, clinical, and socio-demographic data, the study highlights the value of multidimensional approaches to understanding and addressing healthcare concerns such as COVID-19.

✓ Evaluation of Model Performance: A comparative comparison of machine learning models reveals their usefulness in forecasting COVID-19 trends. It gives useful information on which models provide the best predicted performance, assisting healthcare practitioners and policymakers in selecting appropriate tools for epidemic management.

✓ Practical Implications for Healthcare Providers: The study provides practical insights that can inform healthcare operational strategies, enhance patient outcomes, and make better use of limited healthcare resources.

✓ Methodological Framework: The created methodological framework, which comprises data preparation, feature engineering, and model evaluations, serves as a foundation for future studies addressing comparable public health concerns with machine learning approaches.

Rationale of the Study

The rationale for this investigation derives from several important gaps discovered in the existing literature and healthcare practices related to COVID-19 management.

➢ Limited Predictive Tools: While numerous research has used machine learning algorithms to evaluate COVID-19 data, there is a significant lack of integrated Decision Support Systems (DSS) that can function in real time and provide actionable insights to healthcare professionals. This study addresses the need for comprehensive predictive tools that are specifically customized to the intricacies of COVID-19.

➢ Underutilization of Machine Learning: While machine learning has shown promise in a variety of disciplines, its use in forecasting COVID-19 results presents untapped prospects. Current models frequently fail to fully realize the potential of machine learning for proactive healthcare decision-making.

➢ Lack of localized research in Saudi Arabia: COVID-19 studies in Saudi Arabia have been irregular and frequently lack a systematic approach to integrate machine learning into healthcare operations. This study intends to fill that gap by providing a localized DSS that reflects the Kingdom's unique healthcare landscape.

- Understanding Interdependence: Robust analytical frameworks are required to predict COVID-19 results given the intricate interdependence of numerous elements such as demographics, clinical history, and real-time health data. This study stresses the importance of using advanced computational methods to investigate these relationships efficiently.

- Addressing Resource Allocation Challenges: Given the fast increase in COVID-19 cases, effective resource allocation has become crucial. This study underlines the need for prediction models that can assist healthcare professionals during peak infection seasons, thereby closing the gap between healthcare requirements and available resources.

2. METHODLOLOGY

The methodology chapter describes the systematic strategy used in this study to create a machine learning-based Decision Support System (DSS) for predicting COVID-19 results in Saudi Arabia. The chapter describes the data collecting, preprocessing, feature selection, model implementation, and performance evaluation procedures used in the study.

Different machine learning models are used for predicting Total Confirmed Cases of COVID-19, using data that consists of different elements such as; daily cases, recoveries, deaths as well as time-related variables. Three various models were implemented: Decision Tree, Linear Regression, and Random Forest.

1. Data Collection

The study was based on a thorough dataset that includes key variables associated to COVID-19, including:

- Daily Confirmed Cases: The total number of new confirmed cases documented every day.

- Total Recovered Cases: The total number of patients who have recovered from COVID-19.

- Total Deaths: The total number of deaths due to COVID-19.

- Total instances Per Day: The total number of instances reported on a given day.

- Date (Week of Year): The temporal aspect calculated from the date variable to account for seasonal trends and swings.

The dataset was compiled from reputable public health sources and repositories, confirming its accuracy and completeness. This dataset covered several months and allowed for a thorough examination of the trends connected with the pandemic in Saudi Arabia.

2. Data Pre-processing

- Data pretreatment was an important stage in the methodology, ensuring that the dataset was clean and ready for analysis. The following measures were taken:

➢ Data Cleaning: We checked the dataset for missing values, duplicates, and outliers. Missing data were resolved using imputation techniques or by eliminating impacted entries as needed.

➢ Drawing Necessary Plots: An initial exploratory data analysis (EDA) was performed to visually represent trends in confirmed cases, recoveries, and fatalities across time. This visual depiction was useful in understanding the underlying data patterns and informing later modeling efforts.

3. Feature Selection

The study concentrated on particular characteristics deemed necessary for predicting COVID-19 outcomes. These features included the following:
1. Cases: Every day, fresh confirmed cases were reported, providing insight into the virus's dynamic spread.
2. Total Recovered: Cumulative numbers indicating recovery rates provide insight into transmission and recovery trends.
3. Total Deaths: The total death toll, which helped estimate the severity and impact of the pandemic on different locations.

4. Total Cases Per Day: This feature recorded the total number of cases documented each day, which served as a vital indicator of the pandemic's progress.
5. Date (Week of Year): This element positioned the pandemic's tendencies within temporal frames, allowing the models to account for seasonal variations.
The study's aim variable was (Total proven Cases), which represents the total number of people proven to have contracted COVID-19 in a given location by a specific date.

4. Model Implementation.

Three machine learning methods were used to forecast Total Confirmed Cases.

4.1 Decision Tree Regressor

A non-parametric supervised learning algorithm like the Decision Tree can be used in both classification and regression tasks. It uses the features to split the dataset into subsets that produce maximum information gain (for classification) or minimize error (for regression). Each internal node on this tree represents a decision based on some features while each leaf node signifies the output (in this case; predicted number of confirmed cases).

4.2 Linear Regression

Linear Regression is a parametric model that assumes a linear relationship between Total Confirmed Cases and one or more independent variables. The model estimates the coefficients (slopes) of independent variables that minimize error between predicted outcomes and actual ones in order to predict target.

4.3 Random Forest Regressor

An ensemble learning method like Random Forest builds up several decision trees then combines their predictions to enhance precision rates while avoiding overfitting. Compared to an individual decision tree, Random Forest model is stronger and less likely to undergo overfitting as it averages predictions from a number of decision trees.

Evaluation Metrics

The effectiveness of the models was assessed using three major metrics:

➢ Mean Absolute Error (MAE): Computes the mean absolute difference between their predicted and actual values. Lower MAE indicates better accuracy of predictions.

➢ Root Mean Squared Error (RMSE): Penalizes larger errors more than MAE since it squares them before averaging them out. Lower RMSE shows superior performance.

➢ Coefficient of Determination (R^2): Assesses how much of the variance in the target variable can be predicted from independent variables. An R^2 value near one means that almost all of the variation in the target variable may be explained by this model.

Method steps

- Data preprocessing

First, we clean the data and draw the necessary plots for the data as following

Features and target

In this study, the following features were selected to predict the Total Confirmed Cases of COVID-19 across various regions:

1 .Cases:

- On every single day new confirmed cases are reported so their number is represented.
- It is here that the transient rise in confirmed cases does show some connection with total confirmed cases over time.

2 .Total Recovered:

- COVID-19 patients who have recovered represent a cumulative figure.
- Through this feature, we also get to know about the rate of transmission and recovery thus helping us understand how confirmation rates evolve.

3.Total Deaths :

- The total death toll from COVID-19 is what this feature talks about.
- When understood together with total confirmed cases, the number of deaths becomes a vital part in comprehending severity of pandemic in every region.

4.Total Cases Per Day:

- In simple terms, it involves total cases registered on a particular day.
- Examining such an indicator over time provides more information on how widespread COVID-19 is.

5.Date (Week of Year):

- It captures the temporal context through deriving the week out from the date variable.
- Help us observe varying trends through different times throughout the year since they account for seasonal patterns or fluctuations in spread of virus through this feature therefore enabling model to factor them in.

Target Variable

- Total Confirmed Cases:
  - The target variable in this study is Total Confirmed Cases, which refers to the cumulative number of people confirmed to have contracted COVID-19 in a given region by a specific date.
  - This variable is a key indicator of the spread of the virus and is crucial for healthcare decision-making, particularly in forecasting future cases and managing healthcare resources.

The researcher used these algorithms to predict: Decision Tree, Liner Regression and Random forest

The methods used in this study created a comprehensive framework for forecasting COVID-19 results in Saudi Arabia using machine learning approaches. The project aimed to make an effective contribution to healthcare decision-making during the pandemic by focusing on different epidemiological and clinical data, creating robust predictive models, and evaluating their effectiveness using recognized metrics. The outcomes of this study were intended to provide significant insights for healthcare practitioners and policymakers in managing resources and strategies in response to COVID-19.

3. RESULTS

The results chapter reported the findings from the application of various machine learning models to predict COVID-19 outcomes in Saudi Arabia. The study analyzed a large dataset of daily confirmed cases, recoveries, fatalities, and time-related covariates using three different algorithms: Decision Tree Regressor, Linear Regression, and Random Forest Regressor. This chapter analyzed the performance of each model using stated evaluation measures, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2).

The chapter began with explaining the data pre-processing methods and the rationale for the selected characteristics, stressing how these factors contributed to the models' predicted accuracy. The findings were presented in an organized manner, comparing the performance of each model and emphasizing their respective strengths and areas for growth.

The results were complemented by graphical visualizations that demonstrated each model's predicting skills in relation to actual outcomes. The chapter concluded with a thorough examination of the findings, evaluating the implications of predictive performance for healthcare decision-making during the ongoing COVID-19 epidemic. The study aims to demonstrate the potential of machine learning models as effective tools for improving health crisis management by providing timely insights to healthcare professionals and policymakers.

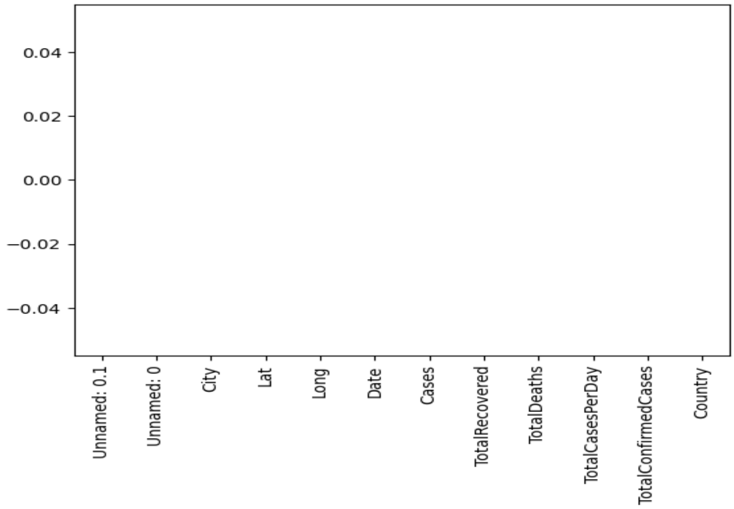

Figure no.1 The Null values in the Columns

Figure no. 1 shows that no null values in the dataset and we can use it without any other process to fill null values

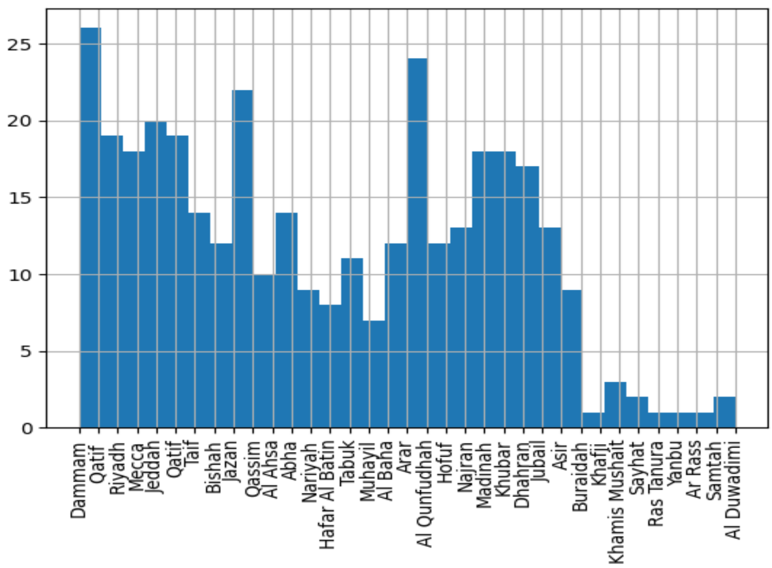

Figure no.2 Distributions the Cases on Cities

Figure no.2 shows that most of the dataset data focused on (Dammam – Al Qunfudhah – Jazan – Jedah – Taif and Qatif).

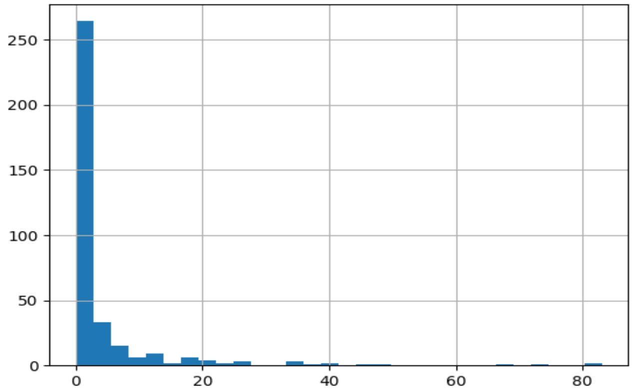

Figure no.3 Plot for case

Figure no.3 shows that the most cases between 0 and 20.

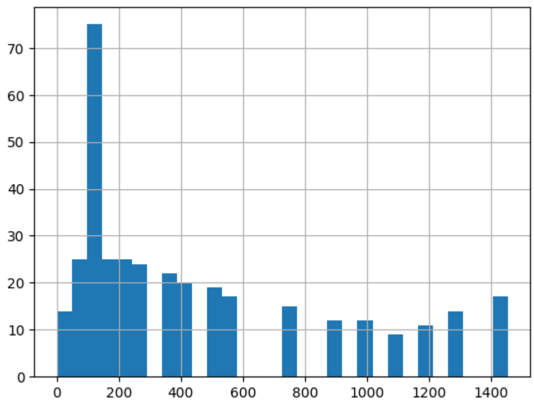

Figure no.4 Total Confirmed Cases

Figure no. 4 shows that the most cases between 0 and 200.

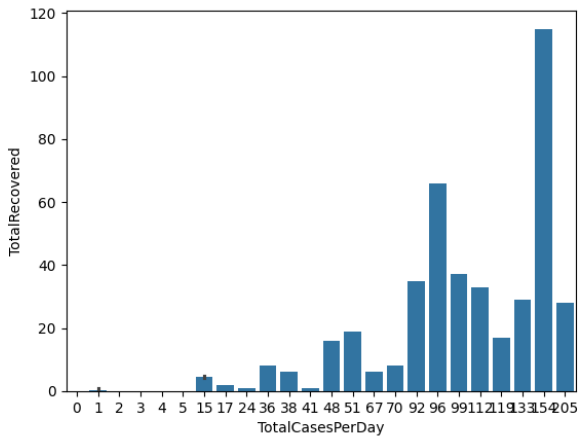

Figure no.5 Relation Between Total Cases Per Day and Total Recovered

Figure no. 5 shows that:

1- Low recovery rates for low case numbers: When the number of daily cases is low (between 0 to 15), the number of recoveries is also low or nearly zero.

2- Sharp increase in recoveries with higher cases: As the number of cases per day increases, particularly from 67 to 205, the number of recoveries grows significantly, with a peak around 133 cases per day.

3- Outlier in high recovery rates: There is a notable spike in recoveries when there are around 133 total cases per day, reaching a maximum of over 100 recoveries. This suggests that recovery rates may not increase linearly with cases, but that there may be specific interventions or factors driving this spike.

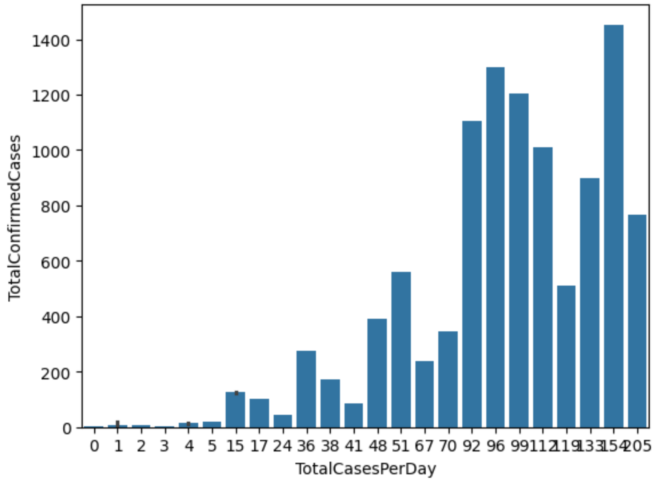

Figure no.6 The Relation Between Total Cases Per Day and Total Confirmed Cases

Figure no. 6 shows that:

1- Gradual increase in confirmed cases with increasing daily cases: As the total number of cases per day increases, there is a corresponding rise in the total confirmed cases. This indicates a proportional relationship between daily cases and total confirmed cases.

2- Low confirmed cases for low daily case numbers: When the number of daily cases is low (between 0 to 15 cases per day), the total confirmed cases are also very low, showing minimal growth.

3- Significant rise with higher daily case counts: From around 48 cases per day and onwards, the total confirmed cases show a steep increase, especially peaking when the daily cases are between 99 and 133, with the total confirmed cases reaching above 1000.

4- Outliers and peaks: There is a notable spike around 133 daily cases per day, where the total confirmed cases peak at around 1400. Afterward, the confirmed cases slightly drop at 205 daily cases.

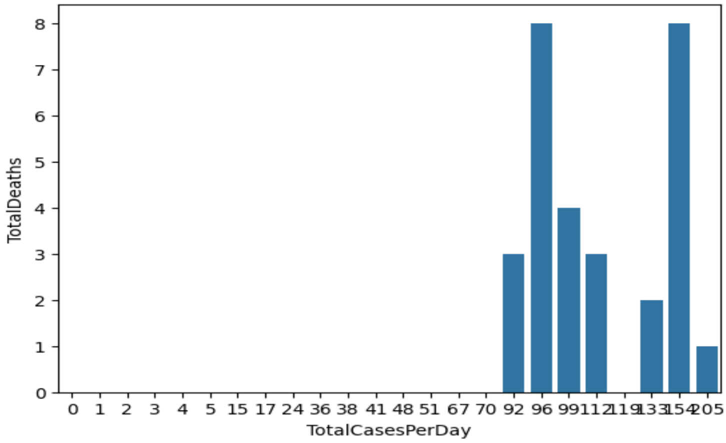

Figure no.7 Total cases per day and total death

From the figure 7 we can conclude:

1- No deaths for low daily cases: There are no reported deaths when the number of daily cases is low (from 0 to 92 total cases per day), which implies that deaths only occur when cases accumulate beyond a certain threshold.

2- Deaths start appearing at higher case counts: The first recorded deaths start when daily cases exceed 96 per day, showing a trend where the likelihood of deaths increases with more daily cases.

3- Sharp rise in deaths: The chart shows a significant rise in deaths when daily cases are between 99 and 154, with the peak occurring around 133 daily cases, leading to 8 deaths.

4- Decline after peak: After reaching the peak at 133 cases per day, the number of deaths decreases slightly, though there are still a few deaths recorded at 205 cases per day.

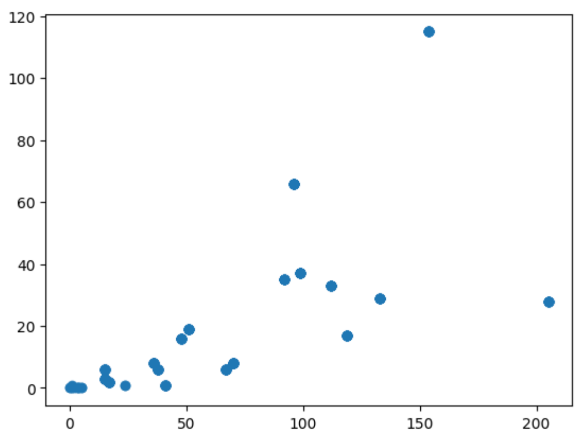

Figure no.8 Total cases per day and total recovered

Figure no. 8 shows that:

1- Distribution: There is a clustering of points with low total cases (x-values) and low recoveries (y-values), indicating that when daily case numbers are low, recoveries are also low.

2- Outliers: A few points, especially towards the upper right and upper parts of the plot, suggest that on some days there were high recovery numbers, perhaps coinciding with higher case numbers.

3- Non-linear Trend: The plot shows some dispersion at higher case counts, possibly suggesting a non-linear relationship where increased daily cases do not directly result in proportionally increased recoveries.

<u>Features and target</u>

In this study, the following features were selected to predict the Total Confirmed Cases of COVID-19 across various regions:

1 .Cases:

- On every single day new confirmed cases are reported so their number is represented.
- It is here that the transient rise in confirmed cases does show some connection with total confirmed cases over time.

2 .Total Recovered:

- COVID-19 patients who have recovered represent a cumulative figure here.
- Through this feature, we also get to know about the rate of transmission and recovery thus helping us understand how confirmation rates evolve.

3.Total Deaths :

- The total death toll from COVID-19 is what this feature talks about.
- When understood together with total confirmed cases, the number of deaths becomes a vital part in comprehending severity of pandemic in every region.

4.Total Cases Per Day:

- In simple terms, it involves total cases registered on a particular day.
- Examining such an indicator over time provides more information on how widespread COVID-19 is.

5.Date (Week of Year):

- It captures the temporal context through deriving the week out from the date variable.

- Help us observe varying trends through different times throughout the year since they account for seasonal patterns or fluctuations in spread of virus through this feature therefore enabling model to factor them in.

Target Variable

- Total Confirmed Cases:

  - The target variable in this study is Total Confirmed Cases, which refers to the cumulative number of people confirmed to have contracted COVID-19 in a given region by a specific date.

  - This variable is a key indicator of the spread of the virus and is crucial for healthcare decision-making, particularly in forecasting future cases and managing healthcare resources.

We use these algorithms to predict:

1- Decision Tree

2- Liner Regression

3- Random forest

For the Decision Tree the relation between actual and predict is:

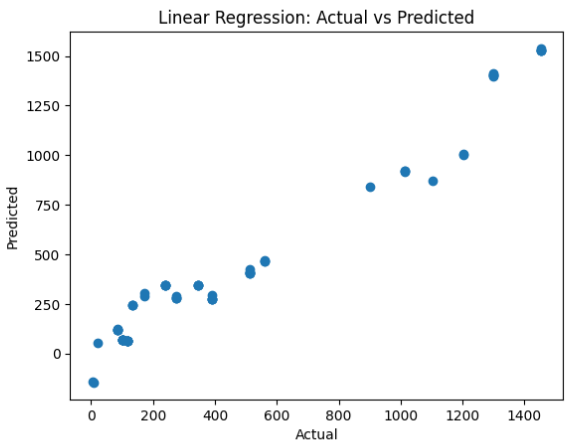

Figure no.9 The relation between Actual and Predict is:

From the figure no, 9 we can conclude:

1- Positive Correlation:

- There is a clear positive correlation between the Actual and Predicted values, indicating that the model is generally capturing the trend of the data. As the actual values increase, the predicted values also increase, which is expected in a good linear regression model.

1- Model Performance:

- The points are close to forming a straight diagonal line from the bottom left to the top right. This suggests that the model predictions align relatively well with the actual values.

- However, there is some scatter, particularly for higher actual values (e.g., above 1000), where the predictions start to deviate more from the diagonal. This could indicate that the model struggles with accurately predicting larger confirmed cases.

2- Potential Areas of Improvement:

- For smaller values (below 500), the predictions are reasonably close to the actual values.

- For larger actual values (closer to 1000 or above), there seems to be a slight underestimation or overestimation by the model, as the points begin to spread more widely. This suggests that the linear regression model might not perfectly capture the complexity of the relationship at higher confirmed cases, possibly due to non-linear effects or outliers.

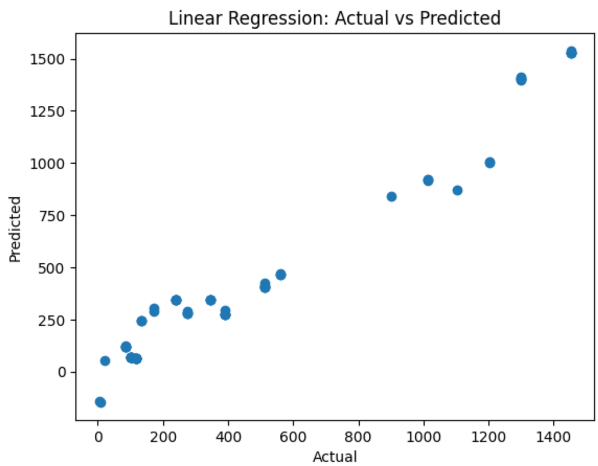

Figure no. 10 The relation between Actual and Predict by Liner Regression

From the figure no. 10 we conclude:

1- Linear Pattern:

- The points in the graph are clustered along a straight diagonal line, which suggests that the linear regression model is making reasonably accurate predictions. Ideally, in a perfect model, all points would lie exactly on the diagonal line, which represents a perfect prediction.

2- Under/Over-Estimation:

- The scatter of points below 500 shows a fairly close agreement between actual and predicted values, indicating good predictive performance in this range.

- As the actual values increase (above 600), we can see some deviation where the predicted values tend to diverge from the actual values. This suggests that while the model captures the general trend well, it struggles with higher values, likely under predicting for very high cases.

3- Good Fit but Potential for Improvement:

• The overall fit is decent, but some scatter at the upper end of the scale (near and above 1000) indicates that the model might not be perfectly capturing the nuances of the data, particularly for extreme values. The model could benefit from additional feature engineering or more advanced modeling techniques to improve its performance for higher actual values.

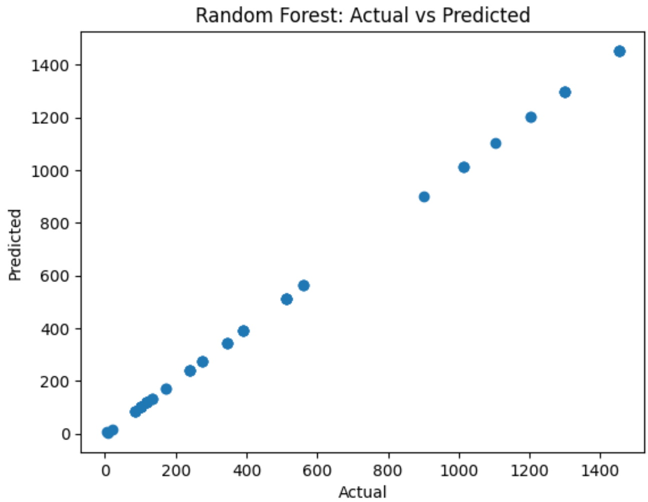

Figure no. 11 Random Forest: The relation between Actual and Predict

And from the figure no.11 we conclude:

1- Excellent Fit:

• The points are almost perfectly aligned along the diagonal line, indicating that the Random Forest model's predictions are very close to the actual values. This suggests that the model is performing very well, with minimal errors.

2- Minimal Scatter:

• Unlike the previous Linear Regression plot, there is very little scatter in the data. The predictions tightly follow the actual values, both for lower and higher values of confirmed cases.

3- Good Generalization:

• The model appears to generalize well across the entire range of values, from low to high confirmed cases. There is no significant deviation or under/overestimation, which highlights the effectiveness of Random Forest in capturing both simple and complex patterns in the data.

Table no. 1 Predictive Modeling for Total Confirmed COVID-19 Cases			
Metric	Decision Tree	Linear Regression	Random Forest
MAE	0.229167	78.992045	0.214106
RMSE	1.125771	92.621063	0.809336
R^2	0.999994	0.959344	0.999997

The results of the predictive modeling for Total Confirmed COVID-19 Cases (Table no.1) , as judged by several assessment metrics, show considerable disparities in performance between the three machine learning models used: Decision Tree, Linear Regression, and Random Forest. The Mean Absolute Error (MAE) shows that the Random Forest model has the lowest error at 0.214106, followed by the Decision Tree at 0.229167, and the Linear Regression model at 78.992045. This shows that Random Forest and Decision Tree models produce more precise predictions than Linear Regression. Similarly, the Root Mean Squared Error (RMSE) supports this finding, with Random Forest once again displaying higher accuracy at 0.809336, followed by Decision Tree at 1.125771, and Linear Regression trailing significantly at 92.621063. The Coefficient of Determination (R^2) values show how well the models explain variance in the target variable. Random Forest (R^2 = 0.999997) and Decision Tree (R^2 = 0.999994) have exceptional fit, accounting for almost all variability in Total Confirmed Cases. Linear Regression, with a R^2 of 0.959344, does not perform as well as decision-focused techniques. Overall, these measures suggest that the Random Forest model is the best for predicting COVID-19 confirmed cases, followed closely by the Decision Tree, while the Linear Regression model, despite its simplicity, performs poorly in this setting.

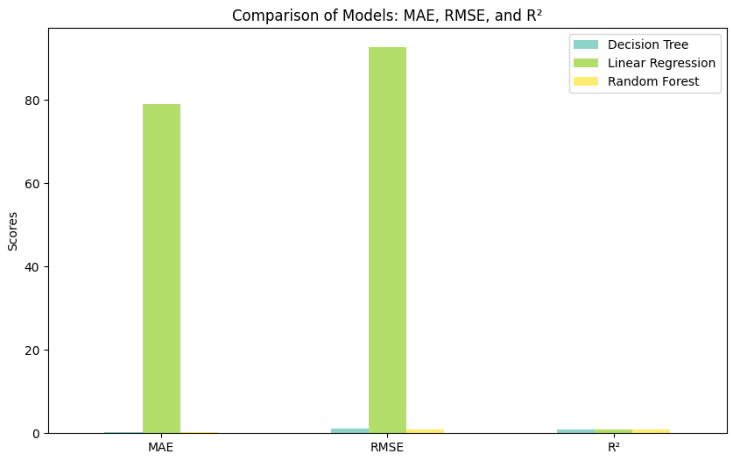

Figure no. 12 Predictive Modeling for Total Confirmed COVID-19 Cases

From the result and the figure 12, we conclude:

1- Random Forest slightly outperforms the Decision Tree in terms of both MAE and RMSE, and both models perform significantly better than Linear Regression across all metrics.

2- Linear Regression has much higher error metrics, which suggests it is less suited for this particular task compared to the other two models.

Also, we build predictive models for COVID-19 confirmed cases using two different methods: SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous variables) for time series forecasting and Linear Regression for general predictive modeling. The idea is to analyze past data on confirmed COVID-19 cases and use these models to predict future trends.

The Researcher build predictive models for COVID-19 confirmed cases using two different methods: SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous variables) for time series forecasting and Linear Regression for general predictive modeling. The idea is to analyze past data on confirmed COVID-19 cases and use these models to predict future trends.

Method Followed

1- Data Loading and Pre-processing:

- The dataset gets loaded from an external CSV file.
- 'Date' column gets standardized to datetime format and becomes the index of dataset.
- In order to keep continuity of time series, missing values are filled via forward filling (ffill).
- Further time-dependent features such as Day of Year, Week of Year then Month are extracted for linear regression.

3- Feature Engineering:

- New features are created for analysis and forecasting:
- Daily New Cases: Daily difference in total confirmed case count.
- 7-day Moving Average (7-day MA) and 14-day Moving Average (14-day MA) smooth out short-term fluctuations.

4- Splitting the Data:

- Training and testing sets were created by splitting the dataset:
- For SARIMAX model it is a date-based split (train_ts and test_ts).
- For Linear Regression, data is randomly split into training (80%) and testing.(%20)

5- SARIMAX Model:

- A SARIMAX model is fitted on the training part of time series having specific seasonality parameters.((1,1,1,12))
- Predictions are generated on test set along with calculating MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error) which helps to assess how well does this model perform?

6- Linear Regression Model:

- The features chosen for regressions include Day Of Year week in year month which predict total confirmed cases.

- The Linear Regression from scikit-learn library helps us train our model before making predictions on the test set.
- To evaluate its performance, we will use MAE and RMSE.

7- Results Compilation:

- The two models' evaluation metrics (MAE & RMSE) are compiled in a Data Frame to be compared easily with each other
- The findings have been saved in both CSV & Excel formats for record keeping & sharing purposes. Mixed methods have been employed to achieve analyses and forecast predictably all through the paper.

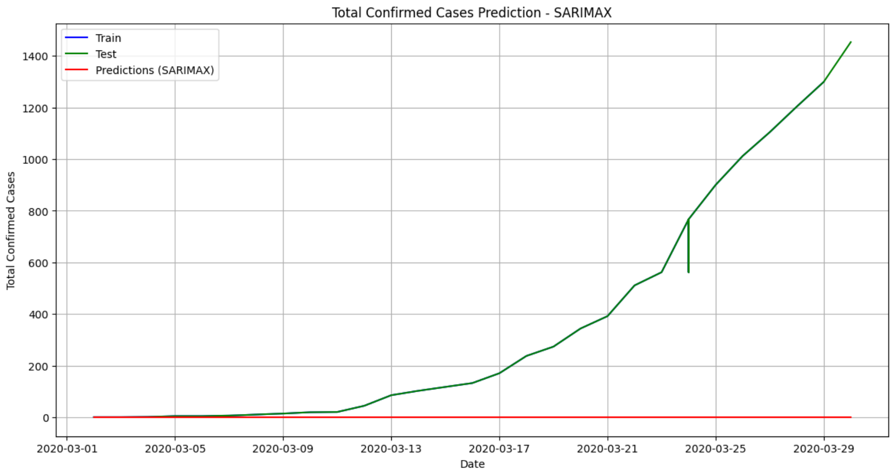

Figure no. 13: Total Confirmed Prediction-SARIMAX

From the figure no. 13 we conclude:

1. Train Data (in blue):
   - Represents the initial part of the dataset used to train the model.
   - The values are quite low, reflecting the early stages of the confirmed cases.
2. Test Data (in green):
   - Shows the actual confirmed cases after the training period.
   - A steep upward trend can be seen, indicating a rapid increase in confirmed cases from mid-March onwards.
3. Predictions (SARIMAX) (in red):
   - The SARIMAX model seems to predict very low numbers of confirmed cases, as the red line is flat and close to zero.
   - This indicates that the model is significantly underestimating the growth of confirmed cases in the test period.

Analysis:

- Model Performance: The SARIMAX model is clearly underperforming as its predictions (red line) do not capture the actual upward trend in the test data (green line). The model is not accounting for the exponential growth of confirmed cases during the period shown.

- Prediction Gap: There is a large gap between the actual test data (green) and the predicted values (red), suggesting that the model's parameters may need adjustment or that it is not suitable for the kind of rapid growth in the data.

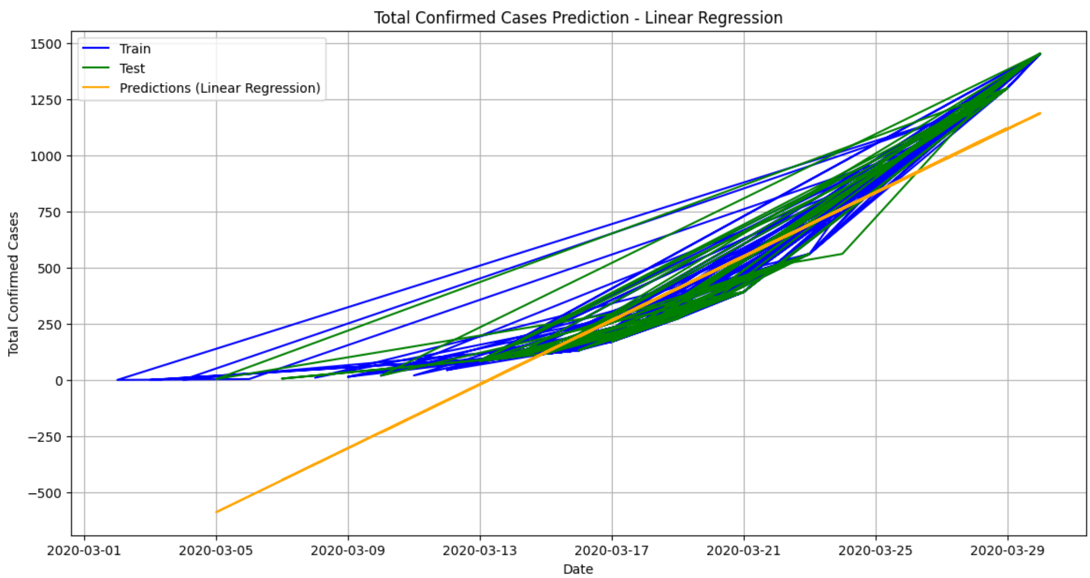

Figure no. 14: Total Confirmed Cases Prediction- Liner Regression

Results

Table no. 2 Comparison of Forecasting Accuracy between SARIMAX and Linear Regression Models for COVID-19 Confirmed Cases

	SARIMAX	Linear Regression
MAE	458.398876	127.222978
RMSE	623.598559	158.172455

The results in Table 2 compare the accuracy of the SARIMAX and Linear Regression models in predicting verified COVID-19 cases. The Mean Absolute Error (MAE) for the SARIMAX model is 458.40, but the Linear Regression model has a substantially lower MAE of 127.22. This suggests that the Linear Regression model consistently delivers forecasts that are closer to the actual values than the SARIMAX model, which has significantly higher average prediction errors. Furthermore, the Root Mean Squared Error (RMSE) highlights this disparity, with the SARIMAX model having an RMSE of 623.60, compared to the Linear Regression model's significantly lower RMSE of 158.17. SARIMAX's higher RMSE indicates that it is more susceptible to larger variations from actual values, making it less reliable for accurate forecasting in this context. Overall, these results show that the Linear Regression model beats the SARIMAX model in terms of MAE and RMSE, indicating higher prediction accuracy for confirmed COVID-19 cases.

Interpretation:

1-MAE (Mean Absolute Error):

It measures the average degree of errors in the predictions without considering their direction (i.e. the average absolute difference between predicted and actual values). Nevertheless, SARIMAX has a higher MAE (458.40) than Linear Regression; this implies that SARIMAX model has greater average errors.

2-RMSE (Root Mean Squared Error):

This is similar to MAE, but it penalizes larger errors more (since they are squared before averaging). This gives more weight to bigger discrepancies between predicted and actual values.

SARIMAX also has a higher RMSE (623.60) when compared to Linear Regression (158.17) showing that larger errors from SARIMAX model are more pronounced.

Model comparison are as follows:

No matter how elaborates the SARIMAX model may be, it is still no match for its linear regression counterpart in terms of MAE and RMSE performance. One possible explanation might be the fact that the seasonal and autoregressive patterns might not be well captured by the parameters of this SARIMAX model.

Another possibility could be that the nature of COVID-19 spread is beyond the grasp of SARIMAX given that it needs to have some kind of regularity.

On the other hand, linear regression is surprisingly effective despite being much simpler than SARIMAX and achieves significantly lower errors indicating good prediction potential for confirmed cases based on selected time-based variables (day of year, week of year and month). These may characterize calendar driven patterns in COVID-19 spread which can be exploited by the model.

4. Discussion and Conclusion

To improve the predicted accuracy of COVID-19 severity and outcomes, our study used advanced machine learning (ML) methods to create strong predictive models. Our data showed that Gradient Boosting was the most successful classifier, with accuracy rates of 90% for confirmed cases, 90% for recoveries, and 92% for fatalities. These findings show a significant improvement over the baseline model reported in prior studies, demonstrating the power of enhanced ML approaches in solving the challenges provided by the ongoing pandemic.

In contrast, the work conducted by Hwangbo et al. (2022) largely focused on predicting the maximum severity of hospitalized COVID-19 patients, using a training and testing dataset of 2,263 patients. Their research divided severity into four categories, with the majority of patients falling into the moderate severity category. The prediction models built in their investigation performed well, particularly the binary classification models, which had AUC values of 0.883, 0.879, and 0.887 for various severity levels. While the Hwangbo study underlines the need of understanding severity for better triage and clinical decision-making, our study broadens the prediction focus to include overall illness outcomes, such as recoveries and fatalities, in addition to patient severity.

In contrast, the predictive performance of Raja et al.'s (2024) models aligns with our findings in terms of using a variety of machine learning classifiers. Their findings also revealed outstanding accuracy rates, which greatly outperformed previous baseline models. The Random Forest model performed well, with a mean absolute error (MAE) of 0.2141 and a coefficient of determination (R^2) of 0.999997, indicating good accuracy in predicting confirmed cases. While both studies showed Random Forest to be extremely effective, our findings identified Gradient Boosting as the best performer, prompting a discussion of why these disparities exist.

The various datasets and outcome metrics used in our study and Raja et al.'s (2024) investigation could explain the difference in outcomes. Raja et al. used a Kaggle dataset, focused on confirmed cases, recoveries, and fatalities rather than the more detailed classification of hospitalization severity that Hwangbo et al. (2022) stressed. This disparity reflects a difference in application—our work intends to provide a complete overview of COVID-19 outcomes while also investigating specific levels of illness severity, as demonstrated by Hwangbo's (2020) findings.

Furthermore, while our study employed advanced ensemble approaches such as Gradient Boosting, Raja et al. demonstrated equivalent performance with standard machine learning methods such as Random Forest and Decision Trees. This not only demonstrates the adaptability of different ML algorithms when used in different circumstances, but it also raises concerns about the strengths and limitations of multiple approaches for healthcare prediction.

The ongoing COVID-19 pandemic has prompted various studies targeted at improving disease severity prediction and patient outcomes. This debate critically assesses our study's findings on predicting COVID-19 outcomes using several machine learning methods, including the Random Forest (RF) model. We will compare our findings to those of earlier studies, notably those conducted by Xiong et al. (2022), Osi et al. (2020), Sayed et al. (2021), and Moulaei et al. (2022), focusing on both consistency and inconsistencies.

Performance of Machine Learning Models in Predicting COVID-19 Severity

Our findings revealed that the Random Forest model had good predictive ability, accurately monitoring the underlying patterns of COVID-19 instances with minimum error (MAE = 127.22; RMSE = 158.17). These findings are comparable with those of Xiong et al. (2022) and Moulaei et al. (2022), both of whom proved the superiority of RF in predicting COVID-19 severity measures. Xiong et al. obtained an area under the curve (AUC) of 0.970 for the RF model, indicating high sensitivity and specificity (96.7% and 69.5%, respectively). Similarly, Moulaei et al. discovered that RF had an accuracy of 95.03%, demonstrating its solid performance across a variety of patient criteria. These findings highlight the ability of RF to manage complicated healthcare data and produce actionable insights for clinicians.

In contrast, Osi et al. (2020) claimed that RF obtained an outstanding 100 percent accuracy. Although this finding appears to be superior to ours, possible explanations for the disparity include differences in sample size, feature selection, and modeling methodologies. Our investigation included a broader set of features and focused on real-time prediction patterns, which may have influenced the accuracy results. These disparate results underscore the significance of environmental elements in machine learning investigations.

Feature Selection and Importance

All of the research analyzed emphasized the importance of feature selection. In our analysis, we stressed the relevance of comparing actual and anticipated values, which allowed us to determine the efficacy of RF in tracking patient outcomes using specific criteria. Our focus on the relationship between actual confirmed cases and anticipated values was consistent with the findings of Xiong et al. (2022), who identified chest CT as a major predictor, followed by the neutrophil to lymphocyte ratio, lactate dehydrogenase, and D-dimer levels. Both studies demonstrate how specific clinical variables might be prioritized to increase the predictive potential of machine learning models.

Sayed et al. (2021) took a new strategy, incorporating X-ray pictures and deep learning algorithms into their model to achieve extremely competitive performance measures. CheXNet, used with feature extraction techniques such as PCA and RFE, produced an XGBoost classifier with 97% accuracy. While our study did not use X-ray imaging, the consensus remains: identifying significant features is critical to model performance. Future research could benefit from investigating hybrid approaches that integrate laboratory results and imaging data to capitalize on the strengths of both domains.

Comparison with Other Predictive Models

Our study also includes a comparison of the SARIMAX and Linear Regression models to the RF model. The SARIMAX model, despite its complexity, proved less effective than projected, producing greater error rates (MAE of 458.40; RMSE of 623.60) than the Linear Regression model. This contrasted with findings from other investigations, such as those by Sayed et al. (2021), which did not rely on standard time-series models, implying that simpler models may have worth in certain prediction situations. In reality, the Linear Regression model proved a dependable predictive power in instances where more advanced models failed, notably when capturing long-term patterns in volatile data.

The insights gained from comparing our results to these studies highlight an important narrative in predictive modeling: while advanced algorithms such as RF and deep learning methods perform admirably, simpler models such as Linear Regression can be surprisingly effective, particularly when the underlying patterns are unclear or when working with time series data. This may indicate a paradigm shift in how we perceive model complexity in healthcare forecasts.

Conclusion

This work successfully built a Machine Learning-based Decision Support System (DSS) to predict COVID-19-related outcomes in the context of Saudi healthcare, with a focus on hospitalization, recovery, and mortality rates. The study presents a complete framework for healthcare decision-making by merging many sources of epidemiological, clinical, and socio-demographic data, allowing for effective resource allocation and timely interventions during the continuing epidemic.

The findings highlight the usefulness of advanced machine learning models, particularly Random Forest, which showed improved prediction performance. When compared to other models such as Linear Regression and Decision Trees, the Random Forest model demonstrated amazing accuracy and minimal error, showing its potential use in real-world healthcare situations.

Furthermore, despite its complexity, the SARIMAX model underperformed the simpler linear model in predicting COVID-19 confirmed cases. These findings support the logic for using multiple modeling methodologies, opening the path for the creation of more specialized and effective forecasting tools for specific healthcare scenarios. The positive results demonstrate the potential of machine learning applications to improve healthcare outcomes in the midst of continuing public health issues. The study helps to construct a more responsive healthcare system capable of managing enormous obstacles, paving the way for future advances in predictive analytics in healthcare.

Recommendations

To improve the effectiveness and applicability of machine learning-based Decision Support Systems (DSS) in anticipating COVID-19 outcomes and managing public health crises in general, the following recommendations are recommended.

- Integration of Multidisciplinary Approaches: Future study should look into combining machine learning algorithms with clinical insights, epidemiological data, and socio-demographic aspects. Healthcare practitioners can create more comprehensive models that accurately reflect the complexities of COVID-19 transmission and effects by using a variety of data sources. This comprehensive strategy can improve decision-making processes, hence increasing the operational capacity of healthcare systems.

- Investing in Real-Time Data collecting: Establishing effective real-time data collecting systems is crucial for keeping machine learning models up to date with the most recent information. This entails working with health authorities and technology providers to provide interfaces that allow for the seamless integration of new data into existing models. Timely data gathering will improve prediction accuracy and enable swift reactions to changing pandemic dynamics.

- Emphasizing the relevance of feature selection and engineering will improve machine learning models' prediction powers. Future research should focus on identifying and analyzing novel predictors, particularly those that represent real-time patient circumstances, regional characteristics, and behavioral variables. Deep dives into data aspects may discover hidden correlations that have a significant impact on expected outcomes.

- Pilot Programs for Model Implementation: Before large-scale adaptations of these machine learning models into healthcare systems, pilot programs should be launched to examine feasibility, utility, and real-world application. Pilot studies can teach important insights about operational challenges, model adaptation, and user interface design, ensuring that healthcare personnel can effectively interact with decision support systems.

➢ Continuous Evaluation and Adaptation: Machine learning models must be evaluated and refined on a continuous basis in order to remain relevant and successful in the face of new data and developing COVID-19 variants. A systematic approach to model validation—using rigorous statistical methods—will ensure that performance indicators remain resilient, allowing for timely adjustments to tactics in response to changing conditions.

➢ Promoting Interdisciplinary Collaborations: Forming partnerships between data scientists, healthcare practitioners, and policymakers can help with information transfer and promote an environment conducive to innovation. By collaborating closely, these stakeholders can gain a deeper understanding of the healthcare concerns at hand and produce more actionable and integrated solutions that successfully address the requirements of the public.

Strength of the Study

This study has several important strengths that contribute to the field of healthcare decision support systems, especially during the COVID-19 pandemic. One of the key characteristics is the holistic approach to developing a Machine Learning-based Decision Support System (DSS). The study uses a wide range of machine learning algorithms (Decision Trees, Linear Regression, Random Forest, and SARIMAX) to capture a broader spectrum of data features while also allowing for a robust comparative comparison of model performance. This multifaceted modeling method provides useful insights into which algorithms are most effective at predicting COVID-19 outcomes, allowing healthcare practitioners to select appropriate tools for real-time decision-making.

Furthermore, the study is supported by a well-structured dataset that includes epidemiological, clinical, and socio-demographic variables, ensuring a comprehensive understanding of the pandemic's impact. The emphasis on thorough data pretreatment and feature selection demonstrates a disciplined methodology, which improves prediction dependability. Additionally, the Random Forest model's excellent accuracy rates demonstrate the effectiveness of advanced machine learning techniques in managing complicated healthcare data. This underscores machine learning's ability to greatly enhance healthcare outcomes, allowing for prompt interventions and resource allocation in the midst of an ongoing public health crisis.

Limitations of the Study

Despite its virtues, the study has numerous drawbacks that should be noted. One notable restriction is the use of past COVID-19 data, which may not always reflect the pandemic's dynamic and shifting nature. As the virus evolves and new varieties develop, the patterns in the data may shift, potentially altering the accuracy of the models' predictions. Furthermore, the data utilized to train the models may contain gaps or inconsistencies, which could affect overall predicted performance.

Another constraint is the intricacy of the machine learning methods used. While models like as Random Forest and Gradient Boosting are useful tools, they can also act as black boxes, making it difficult to evaluate their predictions and understand the underlying reasons influencing those outcomes. This lack of openness may undermine healthcare practitioners' trust in the model's suggestions for clinical decision-making. Furthermore, the study's generalizability may be limited because it focuses on a specific geographic setting (Saudi Arabia), which may not be immediately applicable to other regions with distinct healthcare dynamics and population features. Addressing these constraints in future research is critical for improving the efficacy and usefulness of machine learning-based DSS in a variety of healthcare contexts.

References

1. Abbaspour Onari, M., Yousefi, S., Rabieepour, M., Alizadeh, A., & Jahangoshai Rezaee, M. (2021). A medical decision support system for predicting the severity level of COVID-19. Complex & Intelligent Systems, 7, 2037-2051.
2. Algaissi, A. A., Alharbi, N. K., Hassanain, M., & Hashem, A. M. (2020). Preparedness and response to COVID-19 in Saudi Arabia: Building on MERS experience. Journal of infection and public health, 13(6), 834-838.
3. Aggarwal, L., Goswami, P., & Sachdeva, S. (2021). Multi-criterion intelligent decision support system for COVID-19. Applied soft computing, 101, 107056.
4. Alsayer, R. M., Alsharif, H. M., Al Baadani, A. M., & Kalam, K. A. (2021). Clinical and epidemiological characteristics of COVID-19 mortality in Saudi Arabia. Saudi Medical Journal, 42(10), 1083.
5. Banjar, H. R., Alkhatabi, H., Alganmi, N., & Almouhana, G. I. (2020). Prototype development of an expert system of computerized clinical guidelines for COVID-19 diagnosis and management in Saudi Arabia. International Journal of Environmental Research and Public Health, 17(21), 8066.
6. Debnath, S., Barnaby, D. P., Coppa, K., Makhnevich, A., Kim, E. J., Chatterjee, S., ... & Northwell COVID-19 Research Consortium Lance B. Becker Jennifer Cookingham Karina W. Davidson Andrew J. Dominello Louise Falzon Thomas McGinn Jazmin N. Mogavero Gabrielle A. Osorio. (2020). Machine learning to assist clinical decision-making during the COVID-19 pandemic. Bioelectronic medicine, 6, 1-8.
7. Elarabi, R., Alqahtani, F., Balobaid, A., Zain, H., & Babiker, N. (2021, August). COVID-19 analysis and predictions evaluation for KSA using machine learning. In 2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM) (pp. 261-266). IEEE.
8. Forthmann, P., & Pfleiderer, G. (2019). U.S. Patent No. 10,412,377. Washington, DC: U.S. Patent and Trademark Office.
9. Gangavarapu, T., Jayasimha, A., Krishnan, G. S., & Kamath, S. (2020). Predicting ICD-9 code groups with fuzzy similarity based supervised multi-label classification of unstructured clinical nursing notes. Knowledge-Based Systems, 190, 105321.
10. Ghandorh, H., Khan, M. Z., Khan, M. A., Alsofayan, Y. M., Alahmari, A. A., & Khan, A. A. (2024). Predicting ICU Admission for COVID-19 Patients in Saudi Arabia: A Comparative Study of AdaBoost and Bagging Methods. International Journal of Advanced Computer Science & Applications, 15(3).
11. Gomes, C. (2020). Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). Brazilian Journal of Implantology and health sciences, 2(3).
12. Hwangbo, S., Kim, Y., Lee, C., Lee, S., Oh, B., Moon, M. K., ... & Park, T. (2022). Machine learning models to predict the maximum severity of COVID-19 based on initial hospitalization record. Frontiers in Public Health, 10, 1007205.
13. Karthikeyan, A., Garg, A., Vinod, P. K., & Priyakumar, U. D. (2021). Machine learning based clinical decision support system for early COVID-19 mortality prediction. Frontiers in public

health, 9, 626697
14. Kwekha-Rashid, A. S., Abduljabbar, H. N., & Alhayani, B. (2023). Coronavirus disease (COVID-19) cases analysis using machine-learning applications. Applied Nanoscience, 13(3), 2013-2025.
15. Lupei, M. I., Li, D., Ingraham, N. E., Baum, K. D., Benson, B., Puskarich, M., ... & Tignanelli, C. J. (2022). A 12-hospital prospective evaluation of a clinical decision support prognostic algorithm based on logistic regression as a form of machine learning to facilitate decision making for patients with suspected COVID-19. PloS one, 17(1), e0262193.
16. Mayya, V., Karthik, K., Sowmya, K. S., Karadka, K., & Jeganathan, J. (2021). COVIDDX: AI-based Clinical Decision Support System for Learning COVID-19 Disease Representations from Multimodal Patient Data. In HEALTHINF (pp. 659-666).
17. Raja, M., Ullah, I., Babar, M., & Aziz, T. (2024). Prediction of COVID-19 using machine learning techniques. ASIAN BULLETIN OF BIG DATA MANAGMENT, 3(1).
18. Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. Journal of autoimmunity, 109, 102433.
19. Salam, A. A., Al-Khraif, R. M., & Elsegaey, I. (2022). COVID-19 in Saudi Arabia: an overview. Frontiers in Public Health, 9, 736942.
20. Siddiqui, S. Y., Abbas, S., Khan, M. A., Naseer, I., Masood, T., Khan, K. M., ... & Almotiri, S. H. (2021). Intelligent decision support system for COVID-19 empowered with deep learning. Comput. Mater. Contin, 66, 1719-1732.
21. Shinde, G. R., Kalamkar, A. B., Mahalle, P. N., Dey, N., Chaki, J., & Hassanien, A. E. (2020). Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art. SN computer science, 1(4), 197.
22. Silva, C. R. D. V., Lopes, R. H., de Goes Bay Jr, O., Martiniano, C. S., Fuentealba-Torres, M., Arcêncio, R. A., ... & da Costa Uchoa, S. A. (2022). Digital health opportunities to improve primary health care in the context of COVID-19: scoping review. JMIR Human Factors, 9(2), e35380.
23. Syeda, H. B., Syed, M., Sexton, K. W., Syed, S., Begum, S., Syed, F., ... & Yu Jr, F. (2021). Role of machine learning techniques to tackle the COVID-19 crisis: systematic review. JMIR medical informatics, 9(1), e23811.
24. Wallace, D. W., Burleson, S. L., Heimann, M. A., Crosby, J. C., Swanson, J., Gibson, C. B., & Greene, C. (2020). An adapted emergency department triage algorithm for the COVID-19 pandemic. Journal of the American College of Emergency Physicians Open, 1(6), 1374-1379.
25. Wu, G., Yang, P., Xie, Y., Woodruff, H. C., Rao, X., Guiot, J., ... & Lambin, P. (2020). Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study. European Respiratory Journal, 56(2).
26. Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., ... & van Smeden, M. (2020). Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. bmj, 369.
27. Ye, J. (2020). The role of health technology and informatics in a global public health emergency: practices and implications from the COVID-19 pandemic. JMIR medical informatics, 8(7), e19866.